

## ISOLATION AND STRUCTURAL ANALYSIS OF A GENE CODING FOR A NOVEL TYPE OF ASPARTIC PROTEINASE FROM BUCKWHEAT SEED (*FAGOPYRUM ESCULENTUM* MOENCH)

MIRA DJ. MILISAVLJEVIĆ<sup>1</sup>, GORDANA S. TIMOTIJEVIĆ<sup>1</sup>, SVETLANA R. RADOVIĆ<sup>2</sup>, M. M. KONSTANTINOVIĆ<sup>1</sup>, and VESNA R. MAKSIMOVIĆ<sup>1</sup>

<sup>1</sup>*Institute of Molecular Genetics and Genetic Engineering, 11000 Belgrade, Serbia*

<sup>2</sup>*Faculty of Biology, University of Belgrade, 11000 Belgrade, Serbia*

**Abstract** – A novel type of aspartic proteinase gene was isolated from the cDNA library of developing buckwheat seeds. This cDNA, *FeAPL1*, encoded an AP-like protein lacking the plant-specific insert (PSI) domain characteristic of typical plant aspartic proteinases. In addition the corresponding genomic fragment was isolated. It is demonstrated that this gene does not contain introns. Since bioinformatics analysis of the *Arabidopsis* genome showed that most potential AP genes are intron-less and PSI-less, it appears that “atypical” is an inappropriate word for that class of AP. Isolation of this specific buckwheat gene among the small group of those isolated from other plant species provides a new perspective on the diversity of AP family members in plants.

**Key words:** Aspartic proteinase, buckwheat, cDNA

UDC 633.12 :631.531 :575.1

### INTRODUCTION

Proteolytic enzymes are intricately involved in many aspects of plant physiology and development and therefore are the subject of intensive research interest. As one of the major catalytic classes, aspartic proteinases (EC 3.4.23) are widely distributed, not only in plants but also among vertebrates, yeasts, nematode parasites, fungi, and viruses. According to the MEROPS database (<http://merops.sanger.ac.uk/>), plant APs are distributed among families A1, A3, A11, and A12 of clan AA; and family A22 of clan AD. The majority of plants APs belong to the A1 family, together with pepsin-like enzymes of many origins.

Plant APs have been detected or purified from monocotyledonous and dicotyledonous species as well as gymnosperms (Muller and Gal, 1999) as proteases which are most active at acidic pH, are specifically inhibited by pepstatin A, and contain two aspartic acid residues indispensable for catalytic activity. They are typically distinguished from their non-plant homologs by the presence of a so-called plant-specific insert (PSI), which

is removed during processing and is absent from most mature plant APs. However, there are exceptions to this well-known primary structural organization. These include barley nucellin (Chen and Fould, 1997), its rice ortholog OsAsp1 (Biet al., 2005), *Arabidopsis* CDR1 (Xia et al., 2004), nepenthesins I and II from carnivorous plants (Thud et al., 2004), and tobacco CND41 (Kato et al., 2005). A common feature of this group, often referred to as atypical, AP-like, or a novel class of APs, is absence of the PSI domain. Bioinformatics analysis of the *Arabidopsis* genome sequence revealed 59 AP-like proteins, which provides a new perspective on the diversity of AP family members in plants (Bers et al., 2004).

For the great majority of plant APs, biological functions are still hypothetical and represent a provocative field of investigation. Data related to possible functions come from analysis of the specific expression in certain tissues or under specific conditions, from co-localization studies with putative protein substrates, and from experimental evidence for the processing or degradation of those substrates. Varieties of functions have been pro-

posed for atypical APs. For example, CDR1 is an apoplastic aspartic proteinase, overexpression of which causes resistance to virulent *Pseudomonas syringae* by activating inducible resistance mechanisms (Xia et al., 2004). CND41 is the only DNA-binding aspartic proteinase found in chloroplasts and is involved in degradation of rubisco during leaf senescence (Nakano et al., 1997; Murakami et al., 2000; Kato et al., 2004; Kato et al., 2005). Extracellular nepenthesins have a role in prey digestion (Athauda et al., 2004), while barley nucellin may be involved in nucellar cell death (Chen and Foolad, 1997).

The present paper treats isolation and characterization of cDNA from buckwheat seed encoding for a specific type of aspartic protease. On the basis of its primary structure (characterized by absence of PSI) it should be classified as an AP-like protease. In addition, a genomic fragment containing the corresponding gene and consisting of the coding sequence and the 5' upstream region was also isolated.

## MATERIAL AND METHODS

### *Plant material*

Buckwheat (*Fagopyrum esculentum* Moench, cv. Darja) was field-grown in the greenhouse of the Institute of Molecular Genetics and Genetic Engineering, Belgrade.

### *DNA isolation*

Genomic DNA was isolated as described by Dellaporta et al. (1983). Plasmid DNA was isolated using the Qiagen MiniPrep Kit (Qiagen).

### *Amplification of genomic sequences*

The complete coding sequence of *FeAPLI* was amplified following a 94°C/3 min PCR reaction (94°C/30 s; 60°C/30 s; 68°C/90 s) x 30 cycles and finally post treatment for 10 min at 70°C with Clontech Advantage™ Polymerase using 100 ng of genomic DNA as the template and primers P1 (5'-ATGCCCACTTCTCTCTCTTC-3') and P2 (5'-TTAATTTTTGGATCGATCACATTG-3'). The sequence of primer P1 was derived from the extreme 5'-end and that of P2 from the extreme 3'-end of the coding sequence of *FeAPLI*.

To obtain the 5' regulatory region of the *FeAPLI* gene, we used the modified RACE method (rapid ampli-

fication of cDNA ends) (Chenchik et al., 1995). Adaptor-ligated genomic DNA (100 ng) partially digested with *Eco* RV restriction enzyme (10 U/μg DNA at 37°C for 5 h) was used as the template. For 5' RACE, touchdown PCR amplification (Biometra T1 Thermocycler) was performed in 50 μl of a reaction mixture containing Advantage 2 Polymerase Mix (Clontech), Adaptor primer 1 (Clontech) and gene-specific primers R482 (5'-AGAC-AGTAGGAAAATTTGGTGAGACC-3'). Nested PCR was performed with Nested Adaptor Primer 2 (Clontech) and R470 (5' GGAAAATTTGGTGAGAC-CAAGCTGAG-3') primer. The reaction conditions for both PCRs are described in Brkljacic et al. (2005). The PCR products obtained were subjected to Southern blot analysis using the BioPrime® DNA Labeling System (Gibco BRL) and BluGene® Nonradioactive Nucleic Acid Detection System (Gibco BRL) with *FeAPLI* cDNA as probe. The PCR products that hybridized with the probe were eluted and cloned.

### *Computer-assisted analysis*

Protein sequences were compared using the BLASTP search program (Altschul et al., 1990) and exploring all available sequence databases at the [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) Web server. Sequence analysis was done using the ExPASy program package ([www.expasy.org](http://www.expasy.org)). Sequence alignment was performed with the ClustalW program. The MatInspector professional ([www.genomatix.de](http://www.genomatix.de)), PLACE ([www.dna.affrc.go.jp/PLACE/](http://www.dna.affrc.go.jp/PLACE/)), and PlantCARE (<http://intra.psb.ugent.be:8080/PlantCARE/>) databases were used for prediction of regulatory elements in the 5' regulatory sequence of the *FeAPLI* gene. Prediction of 3D structure of the *FeAPLI* protein was performed using the GeneSilico metaserver (<http://genesilico.pl/meta>) (Kurowski and Bujnicki, 2003).

## RESULTS

We selected and sequenced several clones from the cDNA library of developing buckwheat seed (19-23 DAF stage). Selection was made by subtraction of cDNAs coding for the 13S storage polypeptide (S a m a r d z i c et al., 2004) and metallothionein (Brkljacic et al., 2005), both over-expressed in the defined developmental stage. One of the selected clones ([AY536047](#)) showed amino acid homology of its deduced sequence with a novel class of aspartic proteinases and it was named *FeAPLI*.

The nucleotide sequence of *FeAPLI* consisted of a

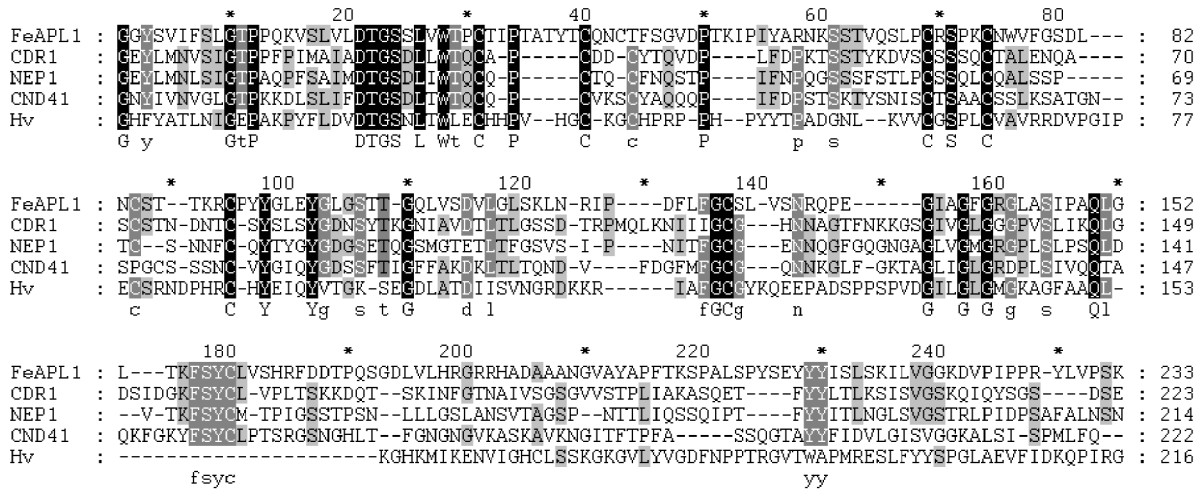


Fig. 1. Amino acid sequence alignment of buckwheat *FeAPL1* and homologs from other plant species. Identical conserved amino acid residues are in black. Similar amino acids are in gray. Fragments of the aa sequence taken for the alignment are notified in slashes. *Fagopyrum esculentum*: *FeAPL1* ([AY536047](#)) /72-447 aa; *Arabidopsis thaliana*: CDR1 ([ACC No. AY243479](#)) /88-437 aa; *Nepenthes gracilis*: NEP1 ([AB114914](#)) /93-437 aa; *Nicotiana tabacum*: CND41 ([D26015](#)) /152-502 aa; *Hordeum vulgare*: nucellin, ([U87148](#)) /36-384 aa

1344 bp long open reading frame (coding for 447 amino acids) and 63 bp and 51 bp of 5' and 3'-UTRs, respectively, followed by a poly(A)-tail. The polypeptide deduced from the *FeAPL1* coding region predicted Mw of 48.6 kDa and an isoelectric point (pI) of 8.78. Analysis of the deduced amino acid sequence of the *FeAPL1* identified the <sup>92</sup>Asp and <sup>313</sup>Asp residues in the DTG/DSG active-site sequence motif characteristic of APs, but unlike most so far identified plant aspartic proteinases, it lacked a plant specific insert (PSI). A hydrophobic signal peptide

in the N-terminal region with a predicted cleavage site between amino acid positions 20 and 21 was identified. Twelve Cys residues, in most cases conserved in atypical plant APs, as well as the so called „flap“ Tyr residue (<sup>168</sup>Tyr corresponding to <sup>75</sup>Tyr in the porcine pepsin numbering) conserved in all animal and plant APs, were also noticed. Four N-glycosylation sites were predicted at positions 113 (NCTF), 130 (NKSS), 154 (NCST), and 357 (NITG). Comparison of the deduced amino acid sequence of *FeAPL1* with the GeneBank and EMBL databases revealed the highest similarity with other known AP-like proteins, such as nepenthesins (NEP I and II), *Arabidopsis* CDR1, tobacco CND41, barley nucellin, and rice *OsAsp1* (Fig. 1). Similarity was also notable with typical APs and with microbial and animal APs, but it was restricted to the active site domains. An NCBI CD search revealed the conserved domain structure, which can be classified as KOG1339, the same as for porcine pepsinogen, cathepsin E, chicken pepsinogen, cow chymosin and candidapepsin 2. Low similarity with atypical APs was clearly evident (Table 1).

Table 1. Similarity between *FeAPL1* and PSI-less aspartic proteinases from other plant species, shown as the percentage of identical amino acids.

FeAPL1	
Gene	% identical aa
NEP I ( <a href="#">AB114914</a> ) <i>Nepenthes gracilis</i>	32
CND41 ( <a href="#">D26015</a> ) <i>Nicotiana tabacum</i>	30
CDR1 ( <a href="#">AY243479</a> ) <i>Arabidopsis thaliana</i>	30
Nucellin ( <a href="#">U87148</a> ) <i>Hordeum vulgare</i>	22
<i>OsAsp1</i> ( <a href="#">AY043217</a> ) <i>Oryza sativa</i>	23

Prediction revealed the tertiary structure of *FeAPL1* (shown in Fig. 2) to be common to those of other enzymes belonging to the aspartic proteinase family whose crystallographic structure has been determined (Kerwin et al., 1999). It consisted largely of β-sheets organized in two lobes divided by a deep active-site cleft, which contained the catalytic aspartates. A six stranded β-sheet covered the bottom of the cleft. A flexible loop

known as the flap, which projected out over the cleft and contained a hydrophobic amino acid and conserved <sup>168</sup>Tyr was also present. Discontinuity of the polypeptide chain in the final model was due to removal of some parts which showed low homology with known APs, thus leading to inaccurate prediction.

The *FaAPLI* cDNA sequence allowed us to develop a strategy for isolation of the corresponding gene from buckwheat genomic DNA. A genomic fragment *gFeAPLI* was obtained by PCR using P1 and P2 primers derived from *FeAPLI* as described in the Material and Methods section. The isolated genomic sequence was identical in length to the cDNA sequence amplified with the same primers, indicating that the *FeAPLI* gene contained no introns. This was confirmed by sequencing.

In order to isolate the 5'-regulatory region of the corresponding gene, a modified 5'-RACE approach was applied using a gene specific primer designed from the 5' end of *FeAPLI*. The longest PCR product was cloned and sequenced. The sequence covered 747 bp overlapping with *FeAPLI* and an additional 1071 bp of the 5'-upstream noncoding region (DO241824). Computer analysis of the 5'-regulatory region predicted three putative TATA boxes at positions 52 bp, 166 bp, and 343 bp relative to the translation start site. Regulatory sequences that could be involved in different responses [hormonal (ABRE, ERE), light (Gbox, RITA1), and pathogen defense (W box)] and elements responsible for seed-specific expression (DOF1, GCN4, SPF1, SEF3, SEF4) were found using overlapping data from three different data-

bases.

## DISCUSSION

Here we report the isolation from buckwheat of full-length cDNA coding for a specific type of aspartic proteases, one that could be relegated to the category of the much less explored plant APs characterized by the lack of a PSI region.

The new class of APs is currently represented by only a few members, but information on this family is increasing. Interestingly, bioinformatics analysis of the *Arabidopsis* genome showed that most of the potential AP genes do not contain a PSI domain (Faro and Gal, 2005). Thus, it appears that "atypical" is an inappropriate word for this class of AP. On the other hand, these remarks raise the question as to why PSI-containing APs are more often discovered in plants if they are not the most abundant. This may be related to their function in highly investigated processes such as seed germination and flowering. It is also possible that they are expressed in greater amounts and thus more easily detected. Another possibility could be that all predicted *Arabidopsis* AP genes do not produce active enzymes or are pseudogenes. According to the MEROPS database, the majority of plant APs belong to the pepsin-like (A1) family. The plant A1 proteases (with or without PSI) from *Arabidopsis* have been annotated and grouped into five distinct subfamilies according to their gene structure and amino acid phylogeny. The largest group, A1-1, consists of mostly intron-less genes. The nucellin-like genes, belonging to group A1-2, have seven introns, while twelve introns are found in group numbers of A1-4, most of which are related to barley phytpepsin. Group A1-4 is the only one whose members contain a PSI domain (Beers et al., 2004). A grouping with similar putative domain organization and active site sequence motifs has recently been described (Faro and Gal, 2005). Thus, intron-less and PSI-less *FeAPLI* should be classified as belonging to group A1-1. Comparison of the deduced amino acid sequence of *FeAPLI* with those of the rest of PSI-less APs showed lower similarity compared to typical plant APs. Clearly, structural similarity is not obviously related to the homology of function, as versatile functions were found for both classes of APs (Simões and Faro, 2004).

The isolated *FeAPLI* cDNA will enable us to produce recombinant protein. This is important as a way of

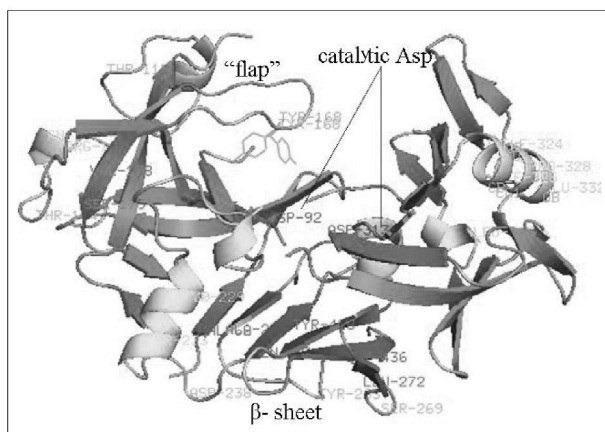


Fig. 2. Tertiary structure of *FeAPLI* protein predicted by molecular modeling with the GeneSilico MetaServer. Catalytic <sup>92</sup>Asp and <sup>313</sup>Asp, conserved <sup>168</sup>Tyr in the flexible loop (flap) and six-stranded  $\beta$ -sheets at the bottom of the cleft are shown.

obtaining information on its biochemical properties, substrate specificity, and mechanism of activation. Also, recombinant protein will serve as a source for producing specific antibodies which can be used to determine protein localization within tissues and colocalization with potential substrates. These results, together with analyses of expression in different physiological processes and those provoked by stress conditions, could be important steps towards clarifying the function of *FeAPLI*. In view of the fact that the functions of plant aspartic proteases are elusive, this will be an important contribution to the resolution of that biological puzzle.

**Abbreviations:** AP - aspartic proteinase; APL - aspartic proteinase-like; PSI - plant specific insert; DAF - days after flowering; RACE - rapid amplification of cDNA ends.

*Acknowledgment* – This work was supported by the Ministry of Science and Environment Protection of the Republic of Serbia (Grant 143017). We thank Dr. Jukka Kervinen for helpful comments and critical reading of the manuscript, as well as for his encouragement to continue our work on aspartic proteinases.

## REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and D.J. Lipman* (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Athauda, S. B., Matsumoto, K., Rajapakshe, S., Kuribayashi, M., Kojima, M., Kubomura, N., Iwamatsu, A., Shibata, C., Inoue, H., and K. Takahashi* (2004). Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases. *Biochem. J.* **381**, 295-306.
- Beers, E.P., Jones, A.M., and A.W. Dickerman* (2004). The S8 serine, C1A cysteine and A1 aspartic protease families in *Arabidopsis*. *Phytochemistry* **65**, 43-58.
- Bi, X., Khush, G.S., and J. Bennett* (2005). The rice nucellin gene ortholog OsAsp1 encodes an active aspartic protease without a plant-specific insert and is strongly expressed in early embryo. *Plant Cell Physiol.* **46**, 87-98.
- Brkljačić, J.M., Majić, D.B., Miljuš-Djukić, J., Bratić, A., Konstantinović, M.M., Milisavljević, M.D., and V.R. Maksimović* (2005). The metallothionein-like gene from buckwheat: structural and functional analysis of the promoter region. *Int. J. Plant Sci.* **166**, 889-900.
- Chen, F., and M. Foolad* (1997). Molecular organization of a gene in barley which encodes a protein similar to aspartic proteinase and its specific expression in nucellar cells during degeneration. *Plant Mol. Biol.* **35**, 821-831.
- Chenchik A., Moqadam F., and P. Siebert* (1995). Marathon cDNA amplification: A new method for cloning full-length cDNAs. *CLONTECHniques X*, 5-8.
- Dellaporta, S.L., Wood, J., and Hicks, J.B.* (1983). A plant DNA mini-preparation: version II. *Plant Mol. Biol. Repts.* **1**, 19-21.
- Faro, C., and S. Gal* (2005). Aspartic proteinase content of the *Arabidopsis* genome. *Curr. Protein Pept. Sci.* **6**, 493-500.
- Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Yokota, A., and F. Sato* (2004). The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. *Planta* **220**, 97-104.
- Kato, Y., Yamamoto, Y., Murakami, S., and F. Sato* (2005). Post-translational regulation of CND41 protease activity in senescent tobacco leaves. *Planta* **222**, 643-651.
- Kervinen, J., Tobin, G.J., Costa, J., Waugh, D.S., Wlodawer, A., and A. Zdanov* (1999). Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO J.* **18**, 3947-3955.
- Kurowski, M.A., and J.M. Bujnicki* (2003). GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.* **31**, 3305-3307.
- Murakami, S., Kondo, Y., Nakano, T., and F. Sato* (2000). Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. *FEBS Lett.* **468**, 15-18.
- Mutlu, A., and S. Gal* (1999). Plant aspartic proteinases: enzymes on the way to a function. *Physiol. Plantarum* **105**, 569-576.
- Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y., and F. Sato* (1997). A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell* **9**, 1673-1682.
- Samardžić, J.T., Milisavljević, M.D., Brkljačić, J.M., Konstantinović, M.M., and V.R. Maksimović* (2004). Characterization and evolutionary relationship of methionine-rich legumin-like protein from buckwheat. *Plant Physiol. Biochem.* **42**, 157-163.
- Simões, I., and C. Faro* (2004). Structure and function of plant aspartic proteinases. *Eur. J. Biochem.* **271**, 2067-2075.
- Xia, Y., Suzuki, H., Borevitz, J., Blount, J., Guo, Z., Patel, K., Dixon, R.A., and C. Lamb* (2004). An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *EMBO J.* **23**, 980-988.

## ИЗОЛОВАЊЕ И АНАЛИЗА СТРУКТУРЕ ГЕНА КОЈИ КОДИРА НОВИ ТИП АСПАРТИЧНЕ ПРОТЕИНАЗЕ СЕМЕНА ХЕЉДЕ (*FAGOPYRUM ESCULENTUM MOENCH*)

МИРА Ђ. МИЛИСАВЉЕВИЋ<sup>1</sup>, ГОРДАНА С. ТИМОТИЈЕВИЋ<sup>1</sup>, СВЕТЛАНА Р. РАДОВИЋ<sup>2</sup>, М. М. КОНСТАНТИНОВИЋ<sup>1</sup>, и ВЕСНА Р. МАКСИМОВИЋ\*<sup>1</sup>

<sup>1</sup>Институт за молекуларну генетику и генетичко инжењерство, 11000 Београд, Србија

<sup>2</sup>Биолошки факултет, Универзитет у Београду, Београд, Србија

Из библиотеке цДНК семена хељде у средњој фази развића изолован је ген који кодира нови тип аспартичне протеиназе. Анализом секвенце ове цДНК (*Fe-APL1*) уочено је одсуство домена карактеристичног само за биљне аспартичне протеиназе, а анализом одговарајућег геномског фрагмента да се ради о гену који не садржи интроне. Биоинформатичким анализа-

ма генома арабидопсиса је показано да је већина потенцијалних гена за аспартичне протеиназе управо са овим особинама, иако је то експериментално доказано само код малог броја гена. Резултати овог рада дају допринос у анализи разноврсности унутар фамилије биљних аспартичних протеиназа.