CROATICA CHEMICA ACTA
CCACAA, ISSN-0011-1643, ISSN-1334-417X

Croat. Chem. Acta **82** (2) (2009) 439–447
CCA-3333

Original Scientific Paper

Arab-1, a GDSL Lipase from the Model Plant, *Arabidopsis thaliana* (L.) Heynh.*

Goran Mikleušević,[a] Branka Salopek-Sondi,[b] and Marija Luić[a,]**

[a]*Department of Physical Chemistry, Ruđer Bošković Institute, P. O. Box 180, 10002 Zagreb, Croatia*
[b]*Department of Molecular Biology, Ruđer Bošković Institute, P. O. Box 180, 10002 Zagreb, Croatia*

RECEIVED SEPTEMBER 1, 2008; REVISED NOVEMBER 6, 2008; ACCEPTED NOVEMBER 7, 2008

Abstract. Bioinformatics analysis of the protein Arab-1 from model plant *Arabidopsis thaliana* (L.) Heynh. was performed at the level of primary, secondary and tertiary structure. Results suggest that Arab-1 belongs to the important but structurally very poorely characterised GDSL (SGNH) family of hydrolases. The amino acids Ser42, Gly118, Asn180, Asp346 and His349 are predicted as the active center residues. According to phylogenetic analysis, Arab-1 is related to *Brassica napus* lipases. Attempts to produce recombinant Arab-1, using pET and pRSET expression systems, and several strains of *Escherichia coli* cells are described; the procedures were found to be inadequate for the production of amounts of protein sufficient for crystallisation. The problems observed in protein production are discussed in view of bioinformatics analyses and calculated physico-chemical parameters.

Keywords: SGNH (GDSL) hydrolase, lipase, *Arabidopsis thaliana*, heterologous protein expression, bioinformatics analysis, phylogenetic analysis

INTRODUCTION

Lipases (triacylglycerol acylhydrolase, EC 3.1.1.3) are enzymes that catalyse the hydrolysis of ester bonds, primarily in neutral lipids such as triacylglycerols. The facts that, in contrast to most other enzymes, they accept a wide range of unnatural substrates and are stable in organic solvents, make them very attractive for industrial applications.[1] A major part of lipases used in today's industrial processes stems from microbial or animal sources. In theory, plant enzymes may have an advantage over animal or microbial enzymes due to their availability from natural sources, lower cost and apparent ease of purification.[2]

To date, very little is known about plant lipases, especially about their 3D-structures and consequently about the mechanism of their catalysis. Based on their 3D-folds, lipases belong to either α/β hydrolase fold enzymes described for the first time by Ollis *et al.*[3] (1992) or to the GDSL family of lipolytic enzymes identified few years later, in 1995 by Upton and Buckley.[4] A canonical α/β hydrolase fold consists of a core of eight mostly parallel β-sheets (not barrel) surrounded by α-helices.[5] All enzymes of this class have a nucleophil-His-acid as a typical catalytic triad. The role of the nucleophil, in lipases, is played by a serine (Ser) residue which is a part of the highly conserved motif Gly-X-Ser-X-Gly (X being any amino acid), located around the middle of the amino acid sequence. Enzymes belonging to the GDSL family share five blocks of highly conserved homology which are important for their classification.[4] The active site serine is located close to the N-terminus.[4] Due to the presence of four strictly conserved residues: Ser, Gly, Asn and His in blocks I, II, III and V, respectively, it has been proposed to change the name of this class of proteins to "SGNH-hydrolases".[6]

To the best of our knowledge there are no published crystal structures for plant lipases. The ultimate goal of our research is, thus, to crystallise and to solve 3D-structure of Arab-1 from *Arabidopsis thaliana* (L.) Heynh., a species widely used as a model organism in plant biology. Due to its relatively small and genetically tractable genome, it has been a subject of intense genetic, biochemical and physiological study for more than 40 years (http://www.arabidopsis.org/). Herein we report the preliminary results of cloning and overex-

* Dedicated to Professor Emeritus Drago Grdenić, Fellow of the Croatian Academy of Sciences and Arts, on the occasion of his 90th birthday.

** Author to whom correspondence should be addressed. (E-mail: marija.luic@irb.hr)

pression, as well as bioinformatics analysis of Arab-1 lipase performed at the level of primary, secondary and tertiary structure.

EXPERIMENTAL SECTION

Cloning, Expression and Purification of Arab-1

Plasmid Construction

The EST clone VBVCD09 containing cDNA encoding the Arab-1 protein was obtained from the Arabidopsis Genomic Resource Center, USA (http://www.arabidopsis.org/). Arab-1 was predicted to be an extracellular protein with a 24 amino acids signal peptide at the N-terminus. Thus Arab-1 cDNA with and without the signal sequence (arab-1sig$^+$ and arab-1sig$^-$, respectively) were amplified by PCR using Pfu DNA polymerase (Invitrogen) and the sets of specific primers listed in Table 1. The reaction conditions for PCR amplification of arab-1sig$^+$ were: 5 min hot start at 368 K, followed by 30 cycles of (1) denaturation at 368 K, for 1 min, (2) annealing at 329 K, for 1 min, and (3) elongation at 345 K, for 1 min. The conditions for arab-1sig$^-$ were the same, except for the annealing temperature which was 328 K. The final elongation reaction for both PCR amplifications was performed for 10 min at 345 K. The PCR products were separated on 1 % agarose gel, extracted from the gel, purified from gel using QIAquick® Gel-extraction kit (QIAGEN) and subcloned into the pGEM-T vector (Promega) as described by manufacturer. The pGEM-T vector containing arab-1 cDNA was isolated from transformants using GenElute™ Plasmid MiniPrep kit (Sigma) and digested with NcoI/XhoI restriction enzymes (Fermentas) for arab-1sig$^+$ and with NdeI/XhoI for arab-1sig$^-$. Both arab-1 constructs were than ligated using T4 DNA ligase (Invitrogen) into expression vector, pET28b (Novagen), which had been digested with NcoI/XhoI and NdeI/XhoI, as well. The newly designed plasmids were named pET28ARAB-Sig$^+$ and pET28ARAB-Sig$^-$.

The pRSET-A expression vector (Invitrogen) was also used to clone and to express Arab-1 protein without signal sequence. PCR amplification of the arab-1sig$^-$ cDNA construct using respective primers was employed to introduce NcoI and XhoI restriction sites at 5'- and 3'-end, respectively. The experimental procedure was the same as described for cloning of the arab-1 cDNA into pET28b vector. The newly designed plasmid was named pRSET-A-ARAB-Sig$^-$. The correct sequences of all cDNA constructs were confirmed by sequence analysis using a ABI PRISM® 3100-Avant Genetic Analyzer (Applied Biosystems) employing the commercial primers T7, forward and reverse.

Heterologous Overexpression and Purification of Arab-1 in Escherichia coli

To overexpress Arab-1, several expression strains of *E. coli* cells (Stratagene) were tested: BL21 (DE3) pLys, BL21 (DE3) Rosetta, and BL21 (DE3) Gold. Constructs containing the Arab-1 gene were introduced into *E. coli* cells and the transformants were grown overnight at 310 K in 10 cm^3 on Luria-Bertrani (LB) medium supplemented with kanamycin (35 μg/cm^3), or ampicillin (50 μg/cm^3) for selection of pET28b and pRSET-A constructs, respectively. Overnight cultures were added into 100 cm^3 of freshly prepared LB medium supplemented with corresponding antibiotics, and proceeded to grow under the same conditions until the optical density, measured at 600 nm, reached 0.5–0.6.

Protein expression was then induced by isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration ranging from 0.4 mmol/dm^3 to 1 mmol/dm^3. The bacteria were subsequently grown for 24 h, in a temperature range of 291 K to 310 K, harvested at 4000 g for 10 min and resuspended in lysis buffer (50 mmol/dm^3 TRIS-HCl, pH = 8.0; 300 mmol/dm^3 NaCl; 10 mmol/dm^3 imidazole). The suspensions were supplemented with lysozyme (1 mg/cm^3), phenylmethanesulfonyl fluoride (PMSF) solution (1 mmol/dm^3), 0.1 % (volume fraction, φ) Triton X-100 and incubated on ice for 30 min. The bacteria were disrupted by freezing the suspensions in liquid nitrogen, followed by thawing (3 cycles). DNase (1 μg/cm^3) was added; the suspension was incubated at room temperature for 30 min, and centrifuged at 14500 g for 30 min to remove cell debris. The supernatant was loaded on 0.4 cm^3 Ni^{2+}-nitrilotriacetic acid agarose column (Qiagen) equilibrated with

Table 1. PCR primers used for amplification of *arab-1* cDNA; the underlined parts of sequences represent restriction sites

Plasmid	Primer	Sequence	Restriction site
pET28b	arab-1sig$^+$ forward	TCCCATGGCTTCTTCACTGAAG	NcoI
	arab-1sig$^+$ reverse	GACCTCGAGTGTATCCACTGTACCAG	XhoI
	arab-1sig$^-$ forward	CGACATATGGAATCTCGATGTAGGCG	NdeI
	arab-1sig$^-$ reverse	GACCTCGAGCTTACTGCTGAAAGAATAC	XhoI
pRSET-A	arab-1sig$^-$ forward	GGACTCGAGGAATCTCGATGTAGGCG	XhoI
	arab-1sig$^-$ reverse	CCCCATGGCTTACTGCTGAAAGAATACCC	NcoI

lysis buffer. The column was washed three times with wash buffer (50 mmol/dm^3 TRIS-HCl, pH = 8.0; 300 mmol/dm^3 NaCl; 20 mmol/dm^3 imidazole) and the bound protein was eluted with elution buffer (50 mmol/dm^3 TRIS-HCl, pH = 8.0; 300 mmol/dm^3 NaCl; 250 mmol/dm^3 imidazole). Protein profiles and purity were checked by SDS-polyacrylamide gel electrophoresis.[7]

Bioinformatics Analysis

Most of the programs and servers used in the prediction of the primary, secondary and tertiary structure of the Arab-1 protein were accessed *via* the Expasy Proteomic Server (http://www.expasy.org/)[8] (Table 2). The sequences of plant GDSL lipolytic enzymes were retrived from Uniprot DataBase[9] by searching for protein sequences similar to Arab-1 using Blastp algorithm.[10] The sequences of SGNH hydrolases with known three-dimensional structures were obtained from the PDB (Protein Data Bank)[11] and SCOP (Structural Clasiffication of Proteins)[12] databases. Protein sequence alignment was performed using ClustalW[13] and pictured by BioEdit Sequence Alignment Editor.[14]

Protein family recognition was accomplished by scanning the protein sequence of Arab-1 against the InterPro protein signature database[15] using InterProScan

Table 2. List of softwares/servers used in bioinformatics analysis of the Arab-1 protein; softwares/servers are listed in alphabetical order

Analysis	Software/Server	Internet addresses
Primary structure analysis	FoldIndex	http://bip.weizmann.ac.il/fldbin/findex
	NetNGly v. 1.0	http://www.cbs.dtu.dk/services/NetNGlyc/
	ProtParam	http://www.expasy.org/cgi-bin/protparam
	SignalP v 3.0	http://www.cbs.dtu.dk/services/SignalP/
	TargetP v. 1.1	http://www.cbs.dtu.dk/services/TargetP/
Secondary structure analysis	DSC	http://www.bmm.icnet.uk/dsc/dsc_form_align.html
	Jpred	http://www.compbio.dundee.ac.uk/~www-jpred/
	Phyre v 0.2	http://www.sbg.bio.ic.ac.uk/~phyre/
	PREDATOR	http://bioweb.pasteur.fr/seqanal/interfaces/predator-simple.html
	PsiPred	http://bioinf.cs.ucl.ac.uk/psipred/
	SSPro	http://scratch.proteomics.ics.uci.edu/
Tertiary structure analysis	DALI	http://www.ebi.ac.uk/dali/
	Phyre v 0.2	http://www.sbg.bio.ic.ac.uk/~phyre/
	PyMol v 0,99rc6	http://pymol.sourceforge.net/
Sequence alignment	BioEdit	http://www.mbio.ncsu.edu/BioEdit/BioEdit.html
	ClustalW	http://www.ebi.ac.uk/Tools/clustalw/
	LALIGN	http://www.ch.embnet.org/software/LALIGN_form.html
Similarity searches	BlastP	http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
Pattern and profile searches	InterProScan	http://www.ebi.ac.uk/InterProScan/
	Pfam HMM Search	http://www.sanger.ac.uk/Software/Pfam/search.shtml
Databases	InterPro	http://www.ebi.ac.uk/interpro/
	PDB	http://www.rcsb.org/pdb/home/home.do
	Pfam	http://www.sanger.ac.uk/Software/Pfam/
	SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
	TAIR	http://www.arabidopsis.org/
	UniProt	http://beta.uniprot.org/

program and against the Pfam database using the Pfam HMM search program.[16] Prediction of secondary structure elements was done by several distinct secondary structure prediction softwares.[17–22]

The Phyre server v 0.2[22] was used to predict the three-dimensional structure of the Arab-1 protein, and visualisation of 3D-structure was accomplished using PyMol v 0,99rc6 (http://pymol.sourceforge.net/). Superposition of the predicted central β-sheet of Arab-1 with the central β-sheet of *Enterococcus faecalis* putative lipase was conducted using the program LSQMAN.[23]

Phylogenetic analysis of plant GDSL lipolytic enzymes was conducted using MEGA package, version 4.0.[24] Two *distance* methods were applied to generate phylogenetic tree, the minimum evolution (ME) and the neighbour-joining (NJ) methods with complete deletion. Poisson correction and p-distance models were used as amino acid supstitution models. The reliability of the phylogenetic tree was checked by bootstrap analysis on 1000 replications.

RESULTS AND DISCUSSION

Cloning, Expression and Purification of Arab-1

The Arab-1 protein has previously been cloned and overexpressed in bacterial cells by Brick *et al*. (1995) and characterised as lipase.[25] We tried to improve the overexpression to obtain sufficient quantities for crystallisation experiments. The expressed sequence tag (EST) clone, VBVCD09 with partial sequence identity to arab-1 cDNA was purchased from the cDNA library of Arabidopsis Genomic Resource Center (TAIR). Sequence analysis revealed complete match of the EST clone to the entire sequence of arab-1 cDNA. A cDNA fragment of 1.2 kb was PCR amplified as described in the Experimental section. Bioinformatics analysis revealed a possible N-terminal signal peptide region in the Arab-1 protein sequence. As this may affect translation in *E. coli*, arab-1 cDNA was constructed with and without its predicted signal peptide leader sequence.

Both constructs of arab-1 cDNA were cloned into the pET28b expression vector. The resulting plasmids, pET28ARAB-Sig$^+$ and pET28ARAB-Sig$^-$, were used for transformation of *E. coli* BL21(DE3)pLys cells. To find optimal expression conditions, different IPTG concentrations (0.4, 0.6, 0.8 and 1 mmol/dm^3), temperatures ranging from 291 K to 310 K, and various times of expression were tested. Very low expression levels were achieved with these conditions, during 24 hours of incubation. We postulated that the differences between codon usages of prokaryotic and eukaryotic organisms could cause low expression level of Arab-1 in *E. coli* cells. Therefore, expression of recombinant Arab-1 protein was attempted in *E. coli* Rosetta strain (Novagen) which contains tRNA unspecific for bacteria which, in turn, can help to acchive higher levels of eukaryotic protein expression. Applying the above conditions, did not improve the level of protein expression. A further alternative plasmid expression system, pRSET-A-ARAB-Sig$^-$, gave the same unsatisfactory result. The quantities of protein obtained after purification by Ni-NTA chromatography were insufficient for biochemical and structural characterisation. Overexpression of eukaryotic proteins, in particular plant lipolytic enzymes, using *E. coli* expression systems is not always straight-forward. Unsuccessful expression in *E. coli* has also been reported for a GDSL esterase from *Alopecurus myosuroides* Huds. and acetylajmaline acetylesterase from *Rauvolfia serpentina* Benth. ex Kurz.[26,27]

Posttranslational modifications such as formation of disulphide bonds, glycosylation and acylations are also important in process of protein synthesis. These processes could interfere with protein folding and drastically reduce the level of expression.[28] We are, thus, considering other expression systems, such as yeast and/or insect-baculoviruses to obtain the amounts of protein required for crystallisation experiments. Furthermore, using plants as bio-factories is a new approach in recombinant protein production. To date experiments reported that transformed plants and/or plant cell cultures are able to synthesize proteins with correct fold, appropriate post translation modification, and activity. Thus this system is promising opportunity for producing of proteins that can not be express in prokaryotic system.[29,30]

Bioinformatics Analysis

A bioinformatics approach was used to characterise the GDSL lipolytic enzyme, Arab-1 from *Arabidopsis thaliana*. The ProtParam program[31] predicts a molecular mass of 42.4 kDa and a pI = 5.32 which attributes acidic character to this protein. The most common amino acid in sequence is leucin, a hydrophobic amino acid with a hydropathy value of 3.800.[30] Likewise, the aliphatic index of 84.84 and the GRAVY (Grand Average of Hidropathicity) index of –0.072 also indicate hydrophobic character.[32] Although Protparam predicts Arab-1 to be a stable protein[33] (instability index is 36.89), the FoldIndex program[34] predictes a large disorder region (87 residues) at the C-terminus which comprises 22.6 % of the sequence. This could interfere with the folding process of Arab-1 and thus lower the expression level.

The program TargetP v. 1.1 predicts Arab-1 to be an extracellular protein with a 24 amino acids signal peptide at the N-terminus (predicted by SignalP).[35] This is the case for several other GDSL enzymes with known

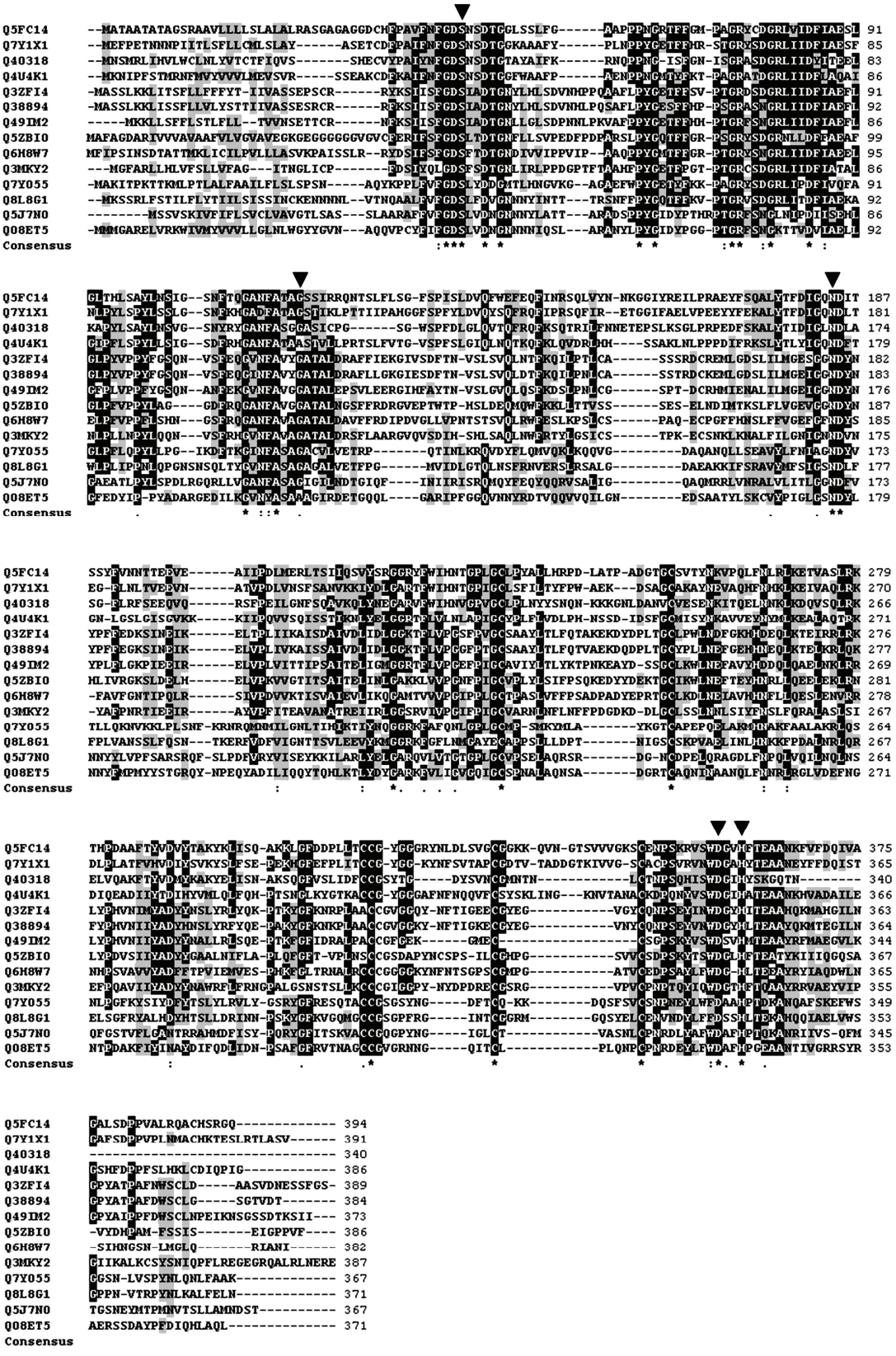

Figure 1. Multiple sequence alignment of GDSL proteins from different plant species. The arrows indicate the amino acid residues constituting the putative active site. Alignment was made using Clustal W and editing was done using the BioEdit Sequence Alignment Editor. Shaded amino acids indicate at least 50 % conservancy.

Table 3. Plant GDSL lipolytic enzymes used for multiple sequence alignment and phylogenetic analysis

Organism	Protein	Uniprot ID
Zea mays L.	acetylcholine esterase	Q5FC14
Hevea brasiliensis (Willd. ex A.Juss.) Müll.Arq.	esterase	Q7Y1X1
Medicago sativa L.	coil protein (alfalfa)	Q40318
Digitalis subalpina Braun-Blanq.	lanatoside 15-*O*-acetylesterase	Q4U4K1
Oryza sativa L.	lanatoside 15'-*O*-acetylesterase-like	Q5ZBI0
Alopecurus myosuroides Huds	esterase precursor	Q6H8W7
Brassica napus L.	lipase 2	Q3ZFI4
Arabidopsis thaliana (L.) Heynh.	Arab-1	Q38894
Brassica napus L.	lipase 1	Q49IM2
Rauvolfia serpentina Benth. ex Kurz	acetylajmalan acetylesterase	Q3MKY2
Chenopodium rubrum L.	GDSL-lipase	Q7Y055
Brassica rapa L. *subsp. pekinensis* (Lour.) Kitam	lipase SIL1	Q8L8G1
Agave americana L.	GDSL-motif lipase	Q5J7N0
Capsicum annuum L.	GDSL-lipase protein	Q08ET5

biochemical functions, such as an esterase from slender foxtail, *Alopecurus myosuroides*,[26] a SGNH hydrolase from the leaf epidermis of *Agave americana* L.,[36] acetylajmaline acetylesterase from *Rauvolfia serpentina*[27] and a GDSL lipase from *Brassica napus* L.[37] All these enzymes are characterised as extracellular proteins playing important roles in developing processes.

Posttranslational modifications such as glycosylation are very common for proteins from eukaryotic organisms. In fact, more than 80 % of all known eukaryotic proteins are glycosylated at asparagine (Asn) residues (N-glycosylation).[28] Glycosylation has a role in protein folding and may affect protein function.[28] There are three possible N-glycosylation sites in Arab-1: Asn105 (sequence: NVSF), Asn138 (sequence: NVSL) and Asn321 (sequence: NFTI) as predicted by NetNGly v. 1.0.[38]

To identify the protein family to which Arab-1 belongs, the databases InterPro and Pfam were scanned using the respective search programs. Both databases described Arab-1 as a member of the "lipase GDSL-SGNH hydrolase" family (InterPro ID: IPR001087, Pfam ID: PF00657).

Searching the UniProt Knowledgebase using Blastp with Arab-1 (Uniprot ID: Q38894) as a query sequence, 500 homologous GDSL protein sequences from plant sources with more than 40 % sequence similarity were found (*E* value < 0.001). Most of them were putative proteins from *Oryza sativa* L. (236) and *Arabi-*

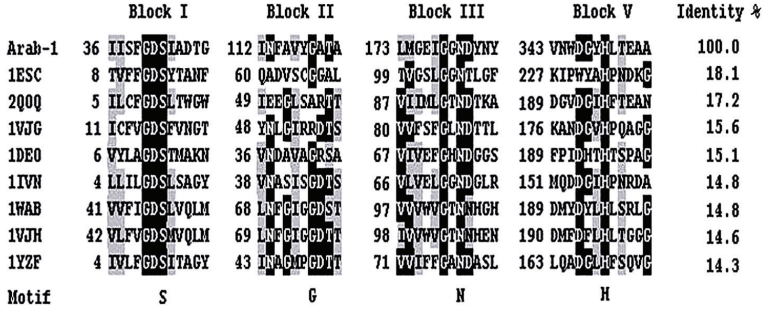

Figure 2. Multiple sequence alignment of four homology blocks of Arab-1 enzyme and SGNH hydrolases with solved crystal structure: *Streptomyces scabies* esterase (PDB code: 1ESC), *Mycobacterium smegmatis* acyltransferase (PDB code: 2Q0Q), N*ostoc sp.* putative lipase (PDB code: 1VJG), *Aspergillus aculeatus* rhamnogalacturonan acetyl-esterase (PDB code: 1DEO), *Escherichia coli* tioesterase I /protease I/ lysophospholipase L1 (PDB code: 1IVN), *Bos taurus* plateled activating factor acetyl hydrolase IBγ subunit (PDB code: 1WAB), *Homo sapiens* plateled activating factor acetylhydrolase IBβ subunit (PDB code: 1VJH), *Enterococcus faecalis* putative lipase(PDB code: 1YZF). SGNH motif is listed below multiple alignment. The sequence identity of Arab-1 to the SGNH hydrolases of known 3D-structure is calculated for the entire sequences lenght using program LALIGN[41] with pairwise sequence alignment.

Figure 3. Secondary structure elements of Arab-1 protein as predicted by several different softwares. The Arab-1 sequence is listed above the prediction. The software packages used in the study are listed at the left side. Secondary structure elements, α helix: cilinder, β-strand: right arrow. Amino acids of potentional active site are enclosed in squares.

dopsis thaliana (105). From different plant species, we selected fourteen GDSL proteins (Table 3) which were not multidomain enzymes and whose functions were biochemically characterised. The sequences were aligned using Clustal W (Figure 1). The results showed several consensus blocks which are very similar to the four conserved sequence blocks of the SGNH hydrolases which contain the amino acids that form enzyme active site and are used to identify the family.[39] To predict active site, the Arab-1 protein sequence was aligned with eight SGNH hydrolases of known three-dimensional structures. Although the overall sequence identity of these proteins was very low, four consensus blocks were identified (Figure 2). Based on similarity of the aligned sequence blocks, we propose that the active site amino acids are Ser42, Gly118, Asn180, Asp346 and His349.

The critical elements of the secondary structure were predicted using several servers (Figure 3). Although there were minor differences, all used software packages have predicted five β-strands and more than 50 % of the structure in loop regions. These are the structural characteristics of the SGNH hydrolases.[6]

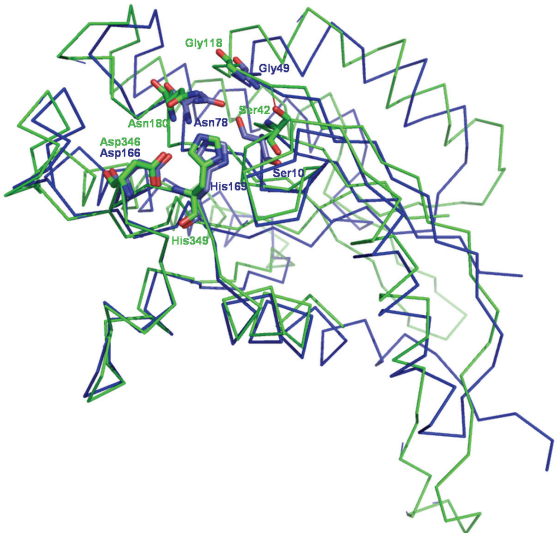

Figure 4. Superpostion of C^{α} backbones of the predicted Arab-1 structure (green) and the 3D-structure of the *E. faecalis* putative lipase, PDB code 1YZF (blue). Active site amino acids are represented by sticks.

To predict three-dimensional structure, the Arab-1 protein sequence was submitted to the Phyre server v 0.2. It predicts protein 3D-structure by scanning the sequence against a non-redundant database of all currently known protein sequences using PSI-BLAST. The 3D- models are than built based on 1D- and 3D-sequence profiles with prediction of secondary structure elements.[22] From the results of the Phyre server, the best model was chosen and processed using the DALI server[40] to find structural homologs. Out of the ten closest homologs (Z score > 13), nine proteins belonged to the SGNH- hydrolase superfamily. Interestingly, even though sequence identity between putative lipase from *Enterococcus faecalis*, PDB code: 1YZF, and Arab-1 was the lowest, compared to other members of SGNH hydrolase superfamily (Figure 2), the DALI server identified the structure of 1YZF to be the closest homolog of Arab-1 (Z score = 32.2, r.m.s.d. = 0.9 Å, and 16 % of sequence identity over the 179 aligned Cα residues). The 3D-model of Arab-1 was visualisated using PyMol v. 099rc6 (see Graphical abstract). As predicted by the Phyre Server, the five β-strands in Arab-1 which compose central β-sheet are: β1 (spans residues 36-39), β4 (spans residues 110-113), β5 (spans residues 172–176), β6 (spans residues 222-224) and β7 (spans residues 285-287). These five β-strands were superimposed with the central β-sheet of the putative lipase from *Enterococcus faecalis* using program LSQMAN.[23] Based on this superposition (r.m.s.d. = 3.038 Å), predicted amino acids of active site in Arab-1 align well with those from *E. faecalis* putative lipase. This is in good accord with the topology preservation of active site amino acids within the family of SGNH hydrolases[6] (Figure 4).

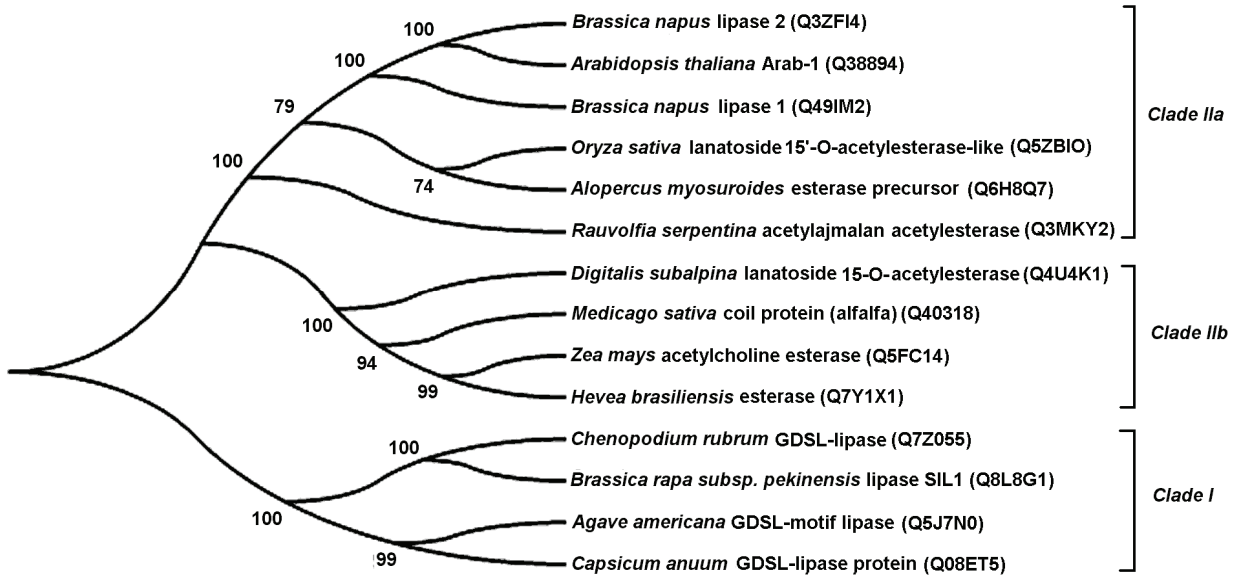

Figure 5. Unrooted neighbour-joining tree of fourteen plant GDSL proteins. Their phylogenetic relationship shows two distinct clades.

Phylogenetic Analysis

A phylogenetic tree comprising fourteen GDSL enzymes from different plant species, constructed using MEGA v. 4.0., exhibited two distinct clades (Figure 5). Clade II contained subgroups, IIa and IIb. Arab-1 belongs to Clade IIa and is the most similar to Lipase 1 and Lipase 2 from *Brassica napus* L.

CONCLUSION

In an attempt to produce sufficient amounts of lipase Arab-1 from *Arabidopsis thaliana* for crystallographic experiments, cloning and overexpression of the protein were undertaken. Two different expression vectors and three different *E. coli* strains were tested, but no significant expression levels were detected. As protein expression in prokaryotes was not successful, expression in the eukaryotic cells will be considered.

Bioinformatics analysis of the target protein was performed at the level of primary, secondary and tertiary structure. Arab-1 was identified as a member of the important, but structurally very poorly characterised GDSL (SGNH) family of hydrolases. The amino acids Ser42, Asp346 and His349 were predicted to represent the catalytic triad residues while Gly118 and Asn180 form the oxyanion hole. Proteins from different biological species with similarity to Arab-1 of 40 % or higher, at the level of primary structure, were analysed phylogenetically. The results indicated two different clades of proteins. Arab-1 showed the highest similarity to the GDSL (SGNH) protein from *Brassica napus*.

Acknowledgements. This work was supported by the Ministry of Science, Education and Sports of the Republic of Croatia (grants: 098-1191344-2943 and 098-0982913-2829). We thank the Arabidopsis Biological Resource Centre, DNA Stock Centre for the EST clone VBVCD09 and Volker Magnus for critical reading of the manuscript.

REFERENCES

1. R. J. Kazlauskas and U. T. Bornscheuer, *Biotransformations with Lipases*, in: D. R. Kelly (Ed.), *Biotechnology*, 2nd ed., *Biotransformations I*, Vol. 8a, VCH Publishers: Weinheim, Germany, 1998, pp. 37–191.
2. P. Villeneuve, *Eur. J. Lipid Sci. Technol.* **105** (2003) 308–317.
3. D. L. Olis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K.H.G. Verschueren , and A. Goldman, *Protein Eng.* **5** (1992) 197–211.
4. C. Upton and J. T. Buckley, *Trends Biochem. Sci.* **20** (1995) 178–179.
5. M. Holmquist, *Curr. Protein Pept. Sci.* **1** (2000) 209–235.
6. A. Mølgaard, S. Kauppinen, and S. Larsen, *Structure* **8** (2000) 373–383.
7. U. K. Laemmli, *Nature* **222** (1970) 680–685.
8. E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, A. Bairoch, *Nucleic Acids Res.* **31** (2003) 3784–3788.
9. A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. S. Yeh, *Nucleic Acids Res.* **33** (2005) D154–159.
10. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *J. Mol. Biol.* **215** (1990) 403–410.
11. H. M. Berman, K. Henrick, and H. Nakamura, *Nat. Struct. Biol.*

10 (2003) 980.
12. A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, *J. Mol. Biol.* **247** (1995) 536–540.
13. D. Higgins, J. Thompson, T. Gibson, J. D. Thompson, D. G. Higgins, and T. J. Gibson, *Nucleic Acids Res.* **22** (1994) 4673–4680.
14. T. A. Hall, *Nucleic Acids Symp. Ser.* **41** (1999) 95–98.
15. R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, E. Birney, M. Biswas, P. Bucher, L. Cerutti, F. Corpet, M. D. R. Croning, R. Durbin, L. Falquet, W. Fleischmann, J. Gouzy, H. Hermjakob, N. Hulo, I. Jonassen, D. Kahn, A. Kanapin, Y. Karavidopoulou, R. Lopez, B. Marx, N. J. Mulder, T. M. Oinn, M. Pagni, F. Servant, C. J. A. Sigrist, and E. M. Zdobnov, *Nucleic Acids Res.* **29** (2001) 37–40.
16. R. D. Finn, J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. L. Sonnhammer, and A. Bateman, *Nucleic Acids Res.* **34** (2006) D247–D251.
17. J. A. Cuff, M. E. Clamp, A. S. Siddiqui, M. Finlay, G. J. Barton, *Bioinformatics* **14** (1998) 892–893.
18. D. T. Jones, *J. Mol. Biol.* **292** (1999) 195–202.
19. J. Cheng, A. Randall, M. Sweredovski, and P. Baldi, *Nucleic Acids Res.* **33** (2005) 72–76.
20. M. Ouali and R. D. King, *Protein Sci.* **9** (2000) 1162–1176.
21. D. Frishman and P. Argos, *Protein Eng.* **9** (1996) 133–142.
22. L. A. Kelley, R. M. MacCallum, and M. J. Sternberg, *J. Mol. Biol.* **299** (2000) 499–520.l
23. G. J. Kleywegt, *Acta Crystallogr.*, *Sect. D* **52** (1996) 842–857.
24. K. Tamura, J. Dudley, M. Nei, and S. Kumar, *Mol. Biol. Evol.* **24** (2007) 1596–1599.
25. D. J. Brick, M. J. Brumlik, J. T. Buckley, J. X. Cao, P. C. Davies, S. Misra, T. J. Tranbarger, and C. Upton, *FEBS Lett.* **377** (1995) 475–480.
26. I. Cummins and R. Edwards, *Plant J.* **39** (2004) 894–904.
27. M. Ruppert, J. Woll, A. Giritch, E. Genady, X. Ma, and J. Stöckigt, *Planta* **222** (2005) 888–898.
28. A. Helenius and M. Aebi, *Annu. Rev. Biochem.* **73** (2004) 1019–1049.
29. M. E. Horn, S. L. Woodard, and J. A. Howard, *Plant Cell Rep.* **22** (2004) 711–720.
30. G. Stefanova, M. Vlahova, and A. Atanassov, *Biol. Plantarum* **52** (2008) 423–428.
31. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, *Protein Identification and Analysis Tools on the ExPASy Server*, in: John M. Walker (Ed.), *The Proteomics Protocols Handbook*, Humana Press, 2005, pp. 571–607.
32. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157** (1982) 105–132.
33. K. Guruprasad, B. V. B Reddy, and M.W. Pandit, *Protein Eng.* **4** (1990) 155–161.
34. J. Prilusky, E. C. Felder, T. Zeev-Ben-Mordehai, E. Rydberg, O. Man, J. S. Beckmann, I. Silman, and J. L. Sussman, *Bioinformatics* **21** (2005) 3435–3438.
35. O. Emanuelsson, S. Brunak, G. von Heijne, and H. Nielsen, *Nat. Protoc.* **2** (2007) 953–971.
36. J. J. Reina, C. Guerrero, and A. Heredia, *J. Exp. Bot.* **58** (2007) 2717–2731.
37. H. Ling, J. Zhao, K. Zuo, C. Qiu, H. Yao, J. Qin, X. Sun, and K. Tang, *J. Biochem. Mol. Biol.* **39** (2006) 297–303.
38. R. Gupta, E. Jung, and S. Brunak, (2004), in preparation.
39. C. C. Akoh, G. C. Lee, Y. C. Liaw, T. H. Huang, and J. F. Shaw, *Prog. Lipid Res.* **43** (2004) 534–552.
40. L. Holm and C. Sander, *J. Mol. Biol.* **233** (1993) 123–138.
41. X. Huang and W. Miller, *Adv. Appl. Math.* **12** (1991) 337–357.

SAŽETAK

Arab-1, GDSL lipaza iz modelne biljke, *Arabidopsis thaliana* (L.) Heynh.

Goran Mikleušević,[a] Branka Salopek-Sondi[b] i Marija Luić[a]

[a]*Zavod za fizičku kemiju, Institut Ruđer Bošković, p. p. 180, 10002 Zagreb, Hrvatska*
[b]*Zavod za molekularnu biologiju, Institut Ruđer Bošković, p. p. 180, 10002 Zagreb, Hrvatska*

Provedena je bioinformatička analiza proteina Arab-1 iz modelne biljke *Arabidopsis thaliana* (L.) Heynh. na razini primarne, sekundarne i tercijarne strukture. Rezultati upućuju na to da Arab-1 pripada važnoj, ali strukturno vrlo slabo istraženoj porodici GDSL (SGNH) hidrolaza. Može se pretpostaviti, da su Ser42, Gly118, Asn180, Asp346 i His349 aminokiseline aktivnog mjesta. Na temelju rezultata filogenetske analize Arab-1 je najsrodniji lipazama iz uljne repice (*Brassica napus* L.). Opisani su pokušaji proizvodnje rekombinantnog Arab-1 proteina uporabom ekspresijskih vektora pET i pRSET te nekoliko sojeva stanica bakterije *Escherichia coli*; pokazalo se, da opisani postupci nisu bili prikladni za dobivanje one količine proteina koja je neophodna za kristalizacijske eksperimente. Problemi priprave proteina diskutirani su i sa stajališta bioinformatičke analize i izračunatih fizičko-kemijskih parametara.