

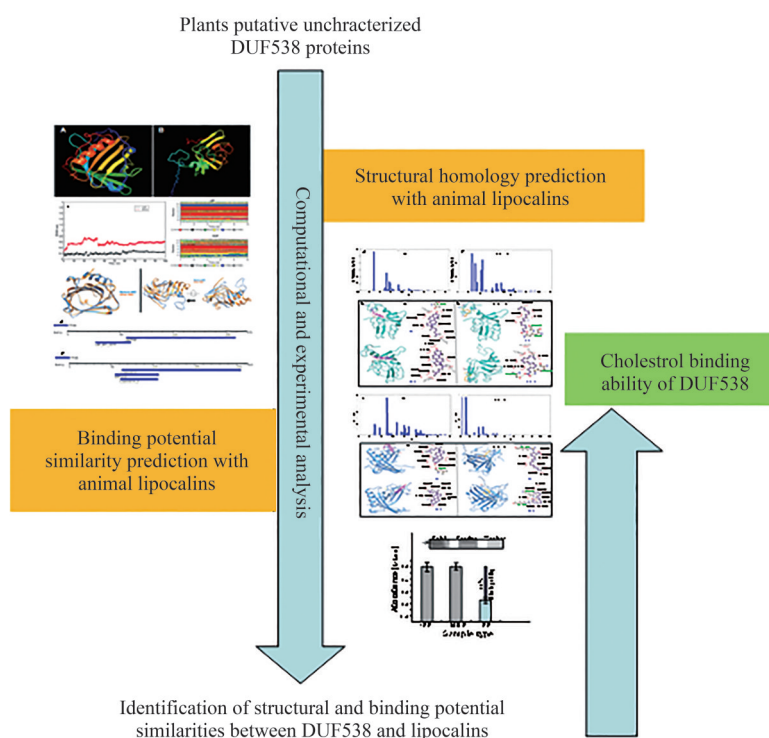


Similarity of Binding Potentials Between Plant DUF538 and Animal Lipocalin: Cholesterol Binding Ability of DUF538*

Ashraf GHOLIZADEH**

(Division of Biochemistry, Department of Animal Sciences, Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran)

Graphical abstract



Abstract Objective DUF538 (domain of unknown function 538) domain containing proteins are known as putative hypothetical proteins in plants. Until yet, there is no much information regarding their structure and function. **Methods** In the present research work, the homologous structures and binding potentials were identified between plant/mammalian lipocalins and plant DUF538 protein by using bioinformatics and experimental tools including molecular dynamics simulation, molecular docking

* This work was supported by a grant from Department of Animal Biology and Research Center for Bioscience and Biotechnology (RCBB), University of Tabriz (6906).

** Corresponding author.

Tel: 98-41-33393924, E-mail: aghz_bioch@yahoo.co.in

Received: May 9, 2022 Accepted: August 30, 2022

and recombinant technology-based techniques. **Results** Molecular docking analysis of their interactions with lipidic ligands including cholesterol and palmitic acid revealed the similar and comparable binding potentials between DUF538 and lipocalin proteins. Both the test proteins were found to have more affinity to cholesterol molecule in compare to palmitic acid. By using recombinant technology-based experiments, the heterologously expressed and purified fused product of DUF538 protein exhibited about 61% cholesterol binding ability. **Conclusion** As a conclusion, plants DUF538 protein family was predicted to be the structural and may be the functional homologues of plants/animals lipocalin superfamily.

Key words binding potential, cholesterol, DUF538, lipocalin, palmitic acid

DOI: 10.16476/j.pibb.2022.0214

DUF538 (domain of unknown function 538) domain containing proteins belong to a large group of uncharacterized protein families as putative hypothetical proteins in plants. They have been found in more than 40 plant species, almost exclusively in *Embryophyta* including the wide ranges of monocotyledonous and dicotyledonous plants, with poorly understood function^[1-3]. Their molecular mass is about 19–21 ku encoding around 170 amino acids. The three-dimensional structures of DUF538 proteins are dominated by β -strands (PDB ID: D1ydua1). DUF538 domain containing proteins or transcripts have often been identified from plants challenged with various abiotic and biotic environmental stresses including drought, elevated temperatures, nutrient deficiency, mixed elicitors, fungal pathogens, nematodes and *Meloidogyne* infection^[1, 3-4]. In general, DUF families, encompassing more than 22% of the entire Pfam database, have no yet clarified function. But based on the recent studies, they might play important roles in different plant stress responses^[3, 5-7]. The involvement of DUF538 proteins in plant growth and developments has also been recently investigated, highlighting the importance of DUF538 protein family in higher plants^[8].

In a different structural and functional studies by our research team, DUF538 proteins have been suggested to be the structural homologues of BPI (bactericidal/permeability increasing) proteins in the innate immune system of mammals^[9]. Their binding potentials to the lipid A moiety of LPS (lipopolysaccharides) on the outer leaflet of the bacterial membranes (as like as BPI of immune system) and lipidic chlorophyll molecules in plants has been later on proposed^[10-11]. Very recently, the

new suggestion of binding capacity of DUF538 proteins towards methylester compounds such as pectin molecules has been reported^[12]. These information led to the conduction of the present research to investigate more about the binding properties of DUF538 domain containing proteins with regards to the lipid/lipid like molecules. By using bioinformatics tools, we found that DUF538 proteins are structurally similar to lipocalins (lipid binding anticalin-like protein family).

Lipocalins are a large and ever expanding group of proteins that exhibit variable structural and functional properties, including lipid binding capacity. The physiological function of lipocalins is usually lies in the transport or storage of hydrophobic compounds such as retinol, vitamins, lipids and steroids^[13-16]. Based on subcellular distribution data analysis, human lipocalins may transfer signals from the extracellular space to the nucleus^[17]. They may also be as biomarkers or metabolic modulators^[16,18]. Lipocalins are mostly secreted proteins and have been characterized from variable prokaryotic and eukaryotic organisms including plant species^[15, 19-20]. Their molecular mass is about 20 ku and their length ranges from 160 to 180 amino acids. The amino acid sequence identity between different lipocalin homologues, with pair wise comparisons, is often falling below 20%. In contrast with their low conservation at the sequence level, they have got conserved folding patterns/three-dimensional structures including eight-stranded anti parallel β -barrel, a short 3_{10} helix and an α -helix at their N-terminus^[21]. This β -dominated structures with flattened or elliptical shape usually encloses an internal ligand-binding site in lipocalin members. The

lipocalins are best known for their binding of a remarkable array of small endogenous and exogenous hydrophobic ligands. A large cup-shaped cavity within the β -barrel and a loop scaffold at its entrance are well adapted for ligand binding. To accommodate ligands of different size and shape, the binding sites of different lipocalins are quite varied^[21-22]. Besides ligand binding ability, based on experimental evidences, lipocalins have been shown to bind to specific cell-surface receptors and internalized by receptor-mediated endocytosis^[21, 23]. In plants, the first evidence for the presence of the putative lipocalin was reported by Bugos *et al.*^[24]. Then the first true plants lipocalins were identified from wheat (*Triticum aestivum*) and *Arabidopsis* (*Arabidopsis thaliana*)^[19]. Finally, using the approaches of data mining of expressed sequence tag (EST) databases, other plant lipocalins with putative functions and cellular locations were identified and characterized^[19].

Considering these together, to increase our knowledge about the structure and function of DUF538 proteins, our attempt was made to identify the similarity of binding potentials of DUF538 and lipocalin proteins with regard to cholesterol and palmitic acid molecules. Using computational and recombinant-based experimental approaches, the lipid-binding ability of DUF538 proteins was investigated and contributed to the novel research challenges of DUF538 protein family in plants. It was speculated that similar to lipocalins, DUF538 proteins may also bind to cholesterol molecules to contribute its trafficking mechanism or metabolizing in plants.

1 Materials and methods

1.1 Computational materials and analysis

The protein sequences of *Arabidopsis* DUF538 (Acc. no. AT5G01610), *Arabidopsis* lipocalin (Acc. no. NP_200615) and *Gorilla gorilla* lipocalin (Acc. no. XP_004048725) were extracted from protein database, NCBI (national center for biotechnology information). The pair wise comparison of protein sequences were carried out by CLASTALW server at <http://www.genome.ad.jp>. Prediction of proteins tertiary structures were performed by Phyre2

internet-based software. For structural stability analysis, proteins were used as input for molecular dynamics simulation performed using GROMACS 2021.5 software. Input structures were prepared with Amberff99SB+ILDN force field. The correct position of hydrogen for all charged amino acids was determined using PROPKA software and the charge of the system was neutralized by adding chlorine ions. The protein was placed in a cubic box using gmx editconf software. The distance between the protein and the walls of the box was considered to be at least 1 nm. Next, TIP3P water molecules were added to the box using the gmx solvate. Energy minimization was performed on the structures using the steepest descent method in order to eliminate steric clashes and inappropriate geometry. In the next step, the system reached equilibrium in NVT (constant temperature and volume) conditions at 298 K in a period of 200 ps in a constant volume using the V-rescale algorithm, and then in NPT conditions at a constant pressure of 1 bar in a period of 200 ps using Parrinello-Rahman algorithm. Finally, the production stage was completed in NPT (constant temperature and pressure) conditions and within 100 ns. van der Waals interactions were calculated with a cut-off equal to 10 Å. Long-range electrostatic interactions were calculated by PME (Particle Mesh Ewald) method. LINCS algorithm was also used to constrain covalent bonds between heavy atoms and hydrogen. All simulations were done in periodic boundary conditions. Finally, the simulation information was stored in 4 ps intervals for analysis. Protein motif analysis was performed by using online motif search tool at <http://www.genome.jp>.

The interaction of the ligands, including cholesterol and palmitic acid, with the target proteins were performed by molecular docking to determine the binding energy and binding sites of the ligands on test proteins. AutoDock version 4.2 software was used for molecular docking performance. PDB for protein and ligand structures were obtained from Rcsb and PubChem databases, respectively. The structures of the test proteins were used as AutoDock input after modifications including heteroatom removal and modeling of deleted protein regions.

After opening the structures with AutoDock software, first the water molecules were selected and removed. Then all the hydrogen atoms were added to the structures and finally all these changes were stored on the structures. Next, a grid boxes were created to select the docking locations on the proteins. For this, the entire protein was placed inside a box. At final step, the proteins as well as the ligand molecules are selected and the Lamarckian Genetic Algorithm (LGA), which determines the number of structural scans, was configured for use at this stage. Docking parameters were also created and the AutoDock program was run and saved all the output conformations as dlz file. After docking, the dlz file was opened and analyzed using AutoDockTools software. The conformations inside this file were ranked based on their binding energy. By using clustering system the binding sites were analyzed.

1.2 Experimental materials

Arabidopsis thaliana plants were prepared from Genetic Engineering Laboratory, Department of Plant Breeding and Biotechnology. Total RNA extraction from the leaves of test plant was carried out by Fermentas Trizol reagent (Cat. no. RN7713C; RNXTM). mRNA purification was done by using QIAGEN mRNA mini preparation kit (Cat. no. 70022). Promega AccessQuickTM RT-PCR System was used for reverse-transcription PCR reaction (Cat. no. A1701). pGEM-T easy plasmid vector for cloning of RT-PCR product was from Promega. Fermentas DNA extraction Kit (Cat. no. K0513) was used for the purification of DNA from the agarose gel. The New England protein fusion and purification system kit (Cat. no. E8000S) utilized for bacterial transformation by *E. coli* strain TB1 and recombinant protein expression by pMALc2X plasmid vector. Materials and columns for the fusion protein isolation and purification and antibody against MBP (maltose-binding protein) were provided in protein fusion and purification system kit. Cholesterol with more than 99% grade was from Sigma (Cat. no. 57-88-5). All the other chemicals used in this research project were of molecular biology grades.

1.3 Cloning and expression of DUF538 as MBP-fused protein

The cloning of DUF538 cDNA from *Arabidopsis*

thaliana leaves was carried out using the following primer pair. Forward: 5'actgaattcgatcagatcttcaacaa-ggttgg3' and Reverse: 5'tctaggatcctcgactcggagaccatttct 3'. In order to do directional cloning in expression vector, *EcoRI* and *BamHI* restriction sites were designed and included at 5' ends of both primers. For RT-PCR based cloning, total RNA was isolated from 0.2 g of leaf material by homogenizing in 2 ml of Trizol reagent. RNA collection was carried out after addition of 200 µl of chloroform and an equal volume of isopropanol, respectively. mRNA population was purified from total RNA sample using oligo dT-columns. According to kit protocol, the RT-PCR reaction was performed with 0.5 µg of mRNA sample in 25 µl of master mix and 1 µl of primer set incubated at 45°C for 45 min and proceeded with PCR cycling. PCR was programmed with 25 cycles as follows: pre-denaturation stage at 95°C for 3 min, (denaturation at 95°C for 1 min, annealing at 56°C for 1.5 min, extension at 72°C for 2 min), and ended with 10 min final extension at 72°C. The amplified products were then extracted from the agarose gel and cloned in pGEM-T easy cloning vector. The cloned fragments proceeded for the partial DNA sequencing in Microsynth center, Switzerland.

For the expression of DUF538 cDNA as MBP-fused product, the *E. coli* fusion protein expression system was used as described in our previous work^[1]. Briefly, the RT-PCR amplification product, after digestion with *EcoRI* and *BamHI* restriction enzymes, ligated into the pMALc2X expression vector and transferred into *E. coli* TB1 competent cells. The transformed cells were plated on LB-agar medium (supplemented with ampicillin and X-gal) at 37°C and a white recombinant clone was selected for the next protein extraction and purification processes.

1.4 Fused protein extraction and purification

In order to extract the fused product from recombinant bacteria, after growing the cells in rich broth medium containing glucose and ampicillin, the expression of fused product was induced by the addition of IPTG at 0.3 mmol/L final concentration for 8 h. Then, the expressing cells after precipitation by centrifugation at 4 000×g for 10 min, dissolved in 25 ml of extraction buffer consisting of 20 mmol/L of Tris-Cl, 200 mmol/L of NaCl, 1 mmol/L of EDTA,

1 mmol/L of sodium azide, and 10 mmol/L of BME (beta mercaptoethanol). After freezing the cells in the same extraction buffer, they were sonicated in short pulses of 15 s and centrifuged at $10\,000\times g$ at 4°C for 20 min. The supernatant was utilized as crude protein extract of fused DUF538 protein. The purification process of fused protein was carried out by using the single-step maltose-binding affinity chromatography. For this purpose, the chromatography column was packed with amylose resin specific for the maltose-binding protein. The bound MBP-fused protein was eluted out from the amylose column by using the specific elution buffer containing crude protein extraction buffer plus 10 mmol/L maltose. The eluted product was analyzed by separating on 10% polyacrylamide gel using SDS-PAGE^[25].

1.5 Western blotting

By using anti-body against maltose-binding protein, the Western analysis was carried out on the induced protein extract. Briefly, the total protein extract was separated on 10% SDS-PAGE and then transferred onto nitrocellulose membrane using blotting buffer containing 0.025 mol/L Tris-Cl pH 8.3, 0.192 mol/L glycine and 20% ethanol. The blot was then kept in TBS-BSA buffer (0.02 mol/L Tris-Cl, pH 7.5, 0.5 mol/L NaCl, 1% BSA) at 4°C and incubated with anti-MAL (maltodextrin-binding protein) antiserum for 2 h at 1 : 500 ratio. After washing with TBS-T (TBS+0.05% Tween 20), the membrane was incubated with alkaline phosphatase conjugated goat-anti rabbit antibody (GAR^{XAP}; diluted 1:20 000 in TBS-BSA buffer) for 1 h. The signal band was visualized using a substrate solution containing 0.33 g/L NBT (nitro blue tetrazolium) and 0.165 g/L BCIP (5-bromo-4-chloro-3-indolyl phosphate) in 0.1 mol/L Tris-Cl buffer containing 0.1 mol/L NaCl and 5 mmol/L MgCl_2 .

1.6 Cholesterol binding ability test of DUF538 fused protein

To analyze the cholesterol binding ability of fused DUF538 protein, about 10 mg/L of purified fused product was dissolved in a buffer consisting of 20 mmol/L of Tris-Cl, 200 mmol/L of NaCl, 1 mmol/L of EDTA and then mixed with amount of cholesterol. The mixture was incubated under laboratory temperature for 20 min. The total cholesterol content

of test sample was assessed by using cholesterol oxidase colorimetric method before and after mixing with fused protein and passing through maltose affinity column. For this, 1 ml of cholesterol assay kit reagent was added to 10 μL of each test sample separately and incubated at 37°C for 10 min to form the pink colored quinone complexes. The absorbance of the samples was then measured at 546 nm and the changes in the absorbance of test samples was estimated and presented as the approximate binding capacity of fused DUF538 towards cholesterol molecules. To test the possible role of MBP in cholesterol binding ability, another control sample including MBP-cholesterol mixture was considered. Each experiment was carried out with three replications. Data points represented on the graphs are the mean \pm SD (standard deviation) of replicates of each test.

2 Results

2.1 Structural similarity analysis

In this study, the protein sequences of *Arabidopsis thaliana* DUF538 (AT5G01610), *Arabidopsis thaliana* lipocalin (NP_200615) and *Gorilla gorilla* lipocalin (XP_004048725) were extracted from NCBI database. Their amino acid sequences were aligned by CLASTALW sequence comparing online server and the obtained results revealed that these proteins are not significantly identical in their primary structures. The homology scores between DUF538 and lipocalin proteins were predicted to be about 20%–22% (Figure 1a). However, the three-dimensional structural analysis by Phyre2 online server indicated that despite primary structures, the tertiary folds do much more support the presence of the highly homologous regions between DUF538 and lipocalin proteins. Comparison of the output data revealed that the folding patterns of these two proteins are confidently dominated by 7–8 anti parallel β -strands with the same organizations and orientations (Figure 1b). The presence of an alpha helical structure at the N-terminus ends is also identical between test proteins, highlighting their three-dimensional structural similarities. The stabilities of 3D structures of test proteins were analyzed by molecular dynamics simulation using

test proteins share the similar pattern of β -plated structures. Comparison of structural conformations before and after molecular dynamics simulations revealed that lipocalin 3D structure is completely stable during simulation time (Figure 2c). But, the 20–25 residues of N-terminus part of DUF538 are structurally changed (Figure 2d). By a survey in protein motif databases using motif search tool at <http://www.genome.jp>, it was clarified that DUF538 and lipocalin proteins don not contain extra motif and their main motifs are DUF538 and lipocalin respectively (Figure 3). Thereby, it is speculated that these motifs may be the substitutes of each other.

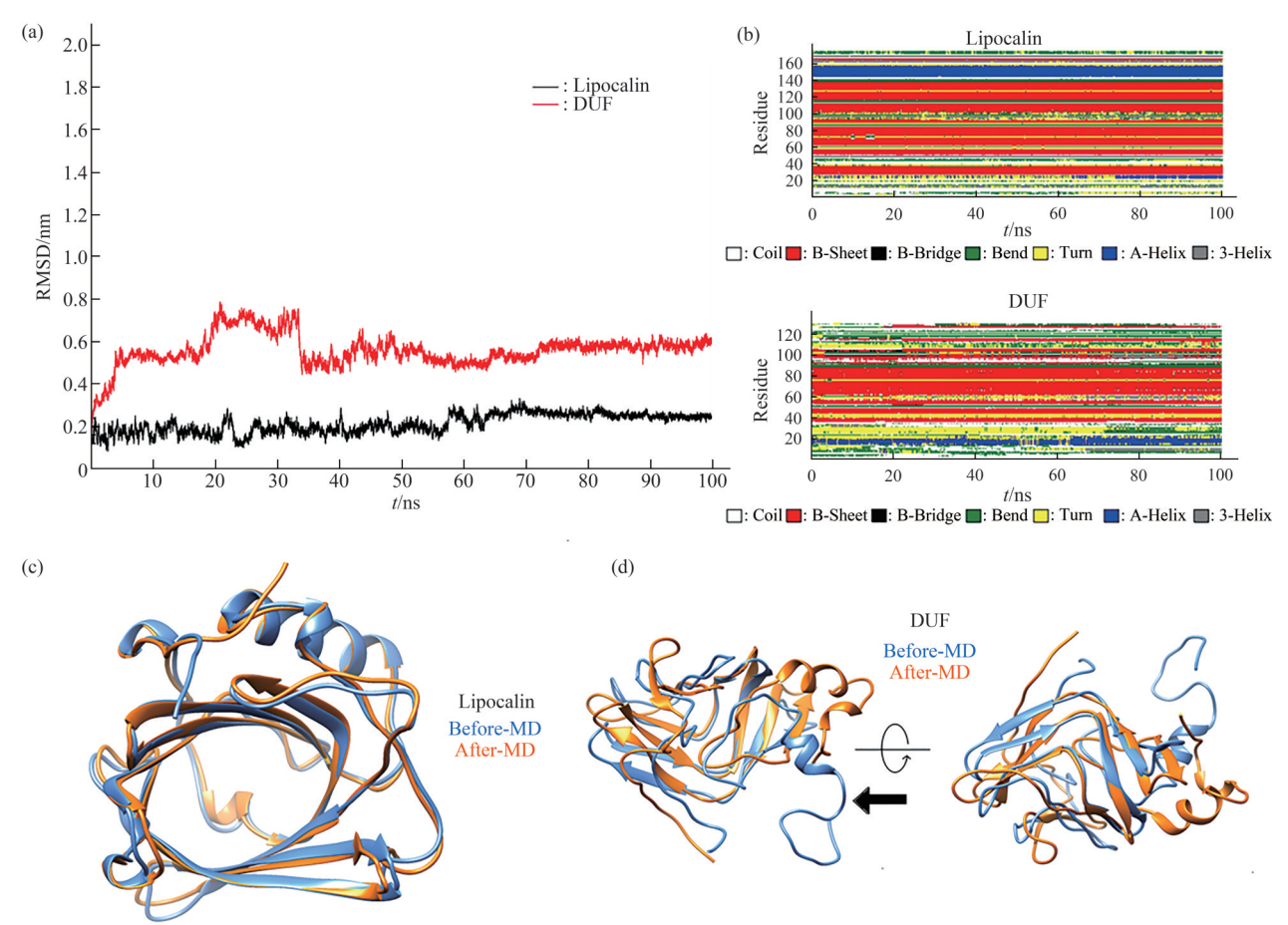


Fig. 2 RMSD, secondary and 3D structural changes during molecular dynamics simulation

(a) RMSD estimation by GROMACS 2021.5 showing the stability of DUF538 and lipocalin structures during simulation period. (b) Secondary structure comparison reveals the similar structural patterns between DUF538 and lipocalin proteins. (c) Conformational comparison showing the conserved structures of lipocalin before (blue) and after (orange) simulation. (d) Conformational comparison showing the conserved structures of DUF538 protein before (blue) and after (orange) simulation.

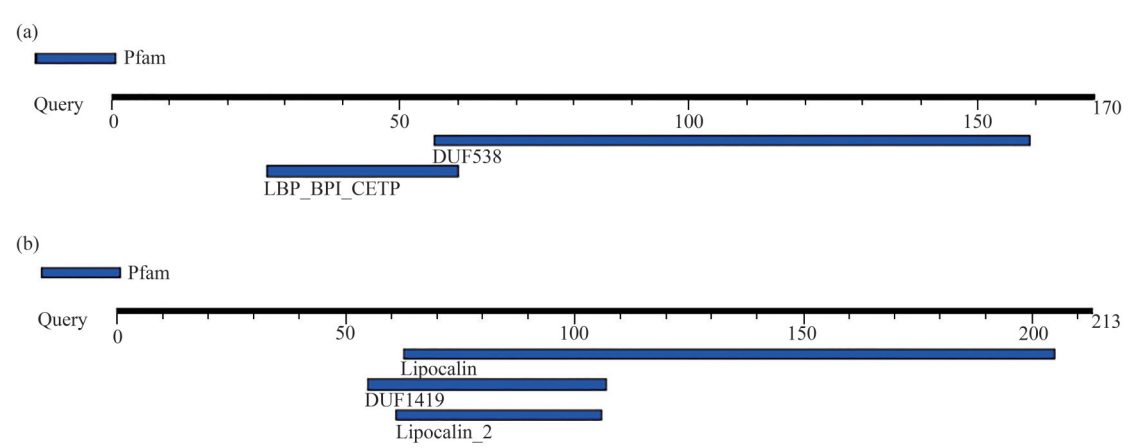


Fig. 3 Protein motif analysis

(a) Motif analysis by using motif search tool at <http://www.genome.jp> revealed that DUF538 protein contains one main motif named DUF538. (b) Motif analysis by using motif search tool at <http://www.genome.jp> detected a main motif named lipocalin on lipocalin protein.

2.2 Binding potentials similarity analysis

In this part of study, AutoDock4.2 and AutoDockTools were used to investigate and analyze the binding of test ligands (cholesterol and palmitic acid) to target proteins including DUF538 and lipocalin. For this, PDB for proteins and ligands structures were obtained from Rcsb and PubChem databases, respectively. Then, using AutoDockTools, polar hydrogens were added and Gasteiger charges were placed on all atoms and rotatable bonds were calculated. Grid box was selected with a distance of 0.375 Å and the target protein was considered inside the box with the dimensions of 126×126×126 and the number of 300 runs. Figure 4 shows the number of clusters, the number of structures in each cluster, and the amount of best binding energy for each cluster in DUF538-cholesterol and DUF538-palmitic acid interactive complexes. According to the obtained results, cluster 1 in the DUF538-cholesterol complex has 113 structures and the binding energy of -8.4 kcal/mol. In the diagram related to DUF538-palmitic acid complex, cluster 1 has 55 members with the binding energy of about -5.4 kcal/mol. Thereby, the results reveals that DUF538 protein has got more potential to bind to cholesterol molecule in compare to palmitic acid. The various energies including electrostatic and van der Waals energies as well as the inhibitory constants of cholesterol and palmitic acid-DUF538 complexes for the first three clusters are shown in Table 1. As compared to DUF-palmitate, DUF-cholesterol complexes show higher number of interactive structures (174 versus 98) and binding energy (24.11 kcal/mol versus 16 kcal/mol). Besides this, the difference between inhibitory constant values of DUF-cholesterol and DUF-palmitic complexes (4.202 μmol/L versus 370.11 μmol/L) is very significant and more considerable. In compare to van der Waals energy, the amount of electrostatic energy related to the binding of cholesterol to DUF538 protein is very low (0.18 versus 6.98), showing the absence of a charged functional group in the structure of the cholesterol molecule. According to the results, comparing the binding potentials (particularly including binding energies and inhibitory constant values) of DUF538 protein to cholesterol and palmitic acid indicates that DUF538 has more affinity to cholesterol molecule than palmitic acid.

In the next step, the binding sites of DUF538 protein to the test ligands were analyzed. Figure 4

shows the binding sites and the surrounding amino acids in two clusters related to the binding of cholesterol and palmitic acid molecules to DUF protein. Chimera and LigPlus softwares were used to prepare 3D and 2D image, respectively. As shown, both the test ligands (including cholesterol and palmitic acid) are mostly bound to the N-terminus part of DUF protein in different complexes. In particular, the results indicated that the conserved N-terminus loop and helical structure of DUF538 protein has decisive role in ligand-protein complex formation (Figure 4). Analysis of interactive amino acids showed that the amino acids Leu87, Pro89, Asn74, Asp71, Trp41, Met50, Lys45, Lys47, Ile46, Asp168, Ile88, Thr106 and Thr107 are involved in DUF-cholesterol complex formation in cluster 1. In this cluster, Thr107 has been shown to interact with hydrogen bonds. In cluster 2, the amino acids Thr106, Thr107, Ile88, Leu87, Pro55, Met50, Asn74, Gly70, Gly48, Asp71, Asp168 and Pro89 are involved in complex formation. In case of DUF-palmitic acid complex, the amino acids Gly48, Lys47, Met50, Ile46, Asn44, Lys45 and Trp41 are involved in complex formation in cluster 1, in which the amino acids Gly48 and Lys47 are interacted with hydrogen bonds. In cluster 2, the amino acids Ser100, Lys126, Trp130, Thr125, Val101, Lys103, Glu93, Lys124 and Leu102 are involved. In this cluster, the amino acids Ser100 and Lys126 are interacted with hydrogen bonds. Analysis of data revealed that cholesterol and palmitic acid have shared the same bound amino acids (including Lys47, Met50, Ile46, Asn44, Lys45 and Trp41) in cluster 1. But, there are no conserved interactive amino acids between DUF complexes in cluster 2. Also data analysis revealed that the bound amino acids of DUF-complexes are not identical between two clusters. In general, analysis of the interactive amino acids in DUF-cholesterol and DUF-palmitic acid complexes showed that the N-terminus part of DUF538 protein is involved in the formation of both complexes and the primary structure of test protein has no much decisive role in ligand-protein interaction.

Similar to DUF538, the molecular docking results of lipocalin protein with cholesterol and palmitic acid molecules have also been investigated. According to the results, cluster 1 in lipocalin-cholesterol complex has 19 interactive structures and the binding energy of -8.3 kcal/mol. While, in the

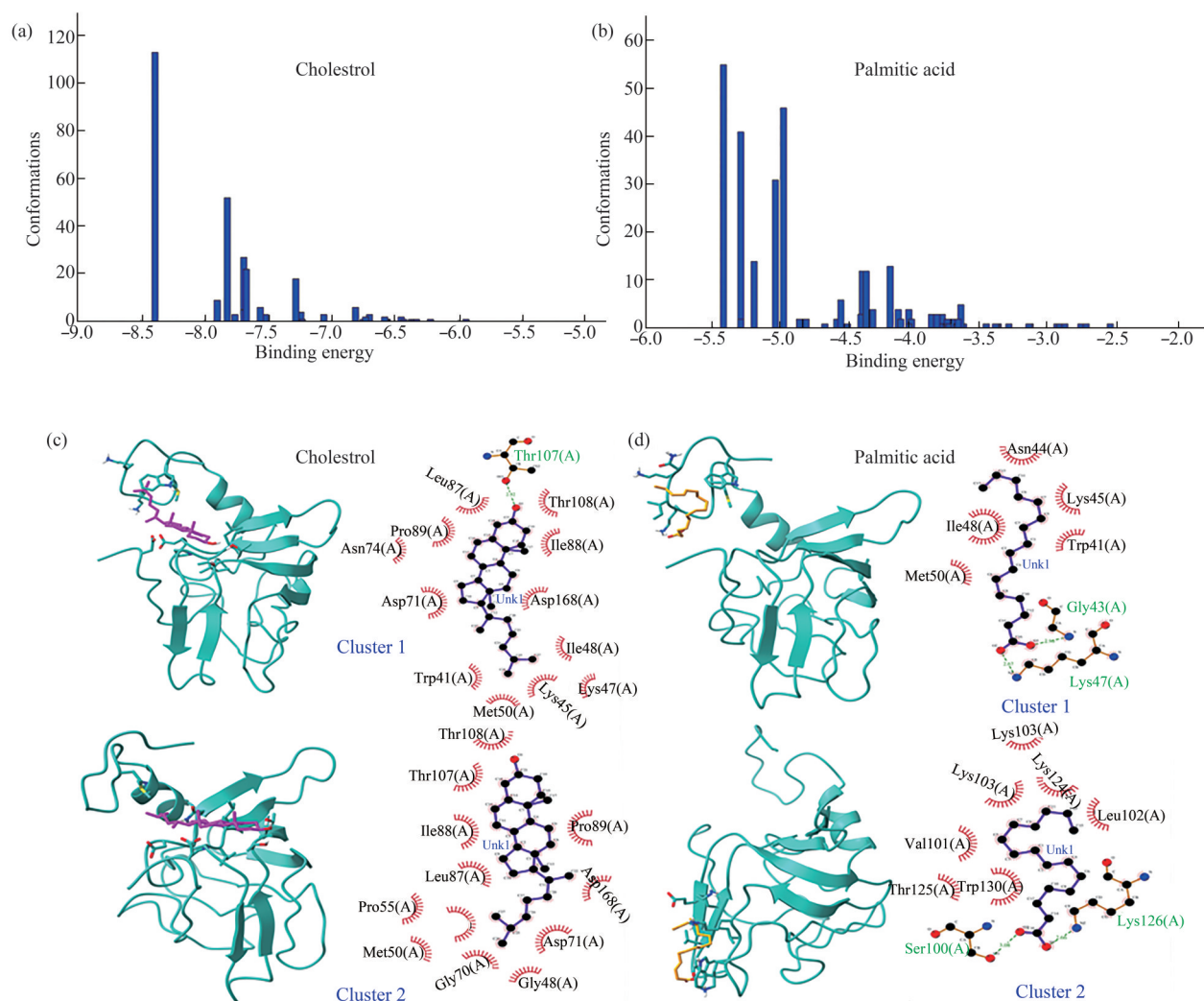


Fig. 4 Binding energies and binding sites analysis of DUF538–ligand complexes

The number of clusters, number of structures and the amount of best binding energy in each cluster of DUF-cholesterol (a) and DUF-palmitic acid (b) complexes were predicted by AutoDock4.2 server. The binding sites of cholesterol (c) and palmitic acid (d) molecules to DUF538 protein in 2 clusters with the highest number of molecular docking members were predicted and compared.

diagram related to lipocalin-palmitic acid complex, cluster 1 has 95 members with the binding energy of -6.99 kcal/mol (Figure 5). Thereby, the results reveal that such as DUF538, lipocalin protein has got more potential to bind to cholesterol molecule in compare to palmitic acid. The energy details and the binding potentials of the top 3 clusters obtained by molecular docking of cholesterol and palmitic acid with lipocalin protein are shown in Table 1. In compare to lipocalin-palmitic acid, the lipocalin-cholesterol complexes have got more binding energy (24.18 kcal/mol versus 20.24 kcal/mol) and low inhibitory constant value (3.961 $\mu\text{mol/L}$ versus

31.11 $\mu\text{mol/L}$). Analysis of electrostatic energies values (0.15 versus 9.95) showed that in compare to van der Waals energies, electrostatic bonds have more decisive role in lipocalin-palmitate complex formation. The overall results revealed that similar to DUF538 protein, lipocalin has also more affinity to cholesterol molecule rather than palmitic acid.

Same to DUF538, the binding sites and the surrounding amino acids of cholesterol and palmitic acid-lipocalin complexes in clusters 1 and 2 were also investigated (Figure 5). In cluster 1 of lipocalin-cholesterol complex, the amino acids Leu77, Val202, Gln84, Phe203, Pro204, Val205, Pro206, Asp80,

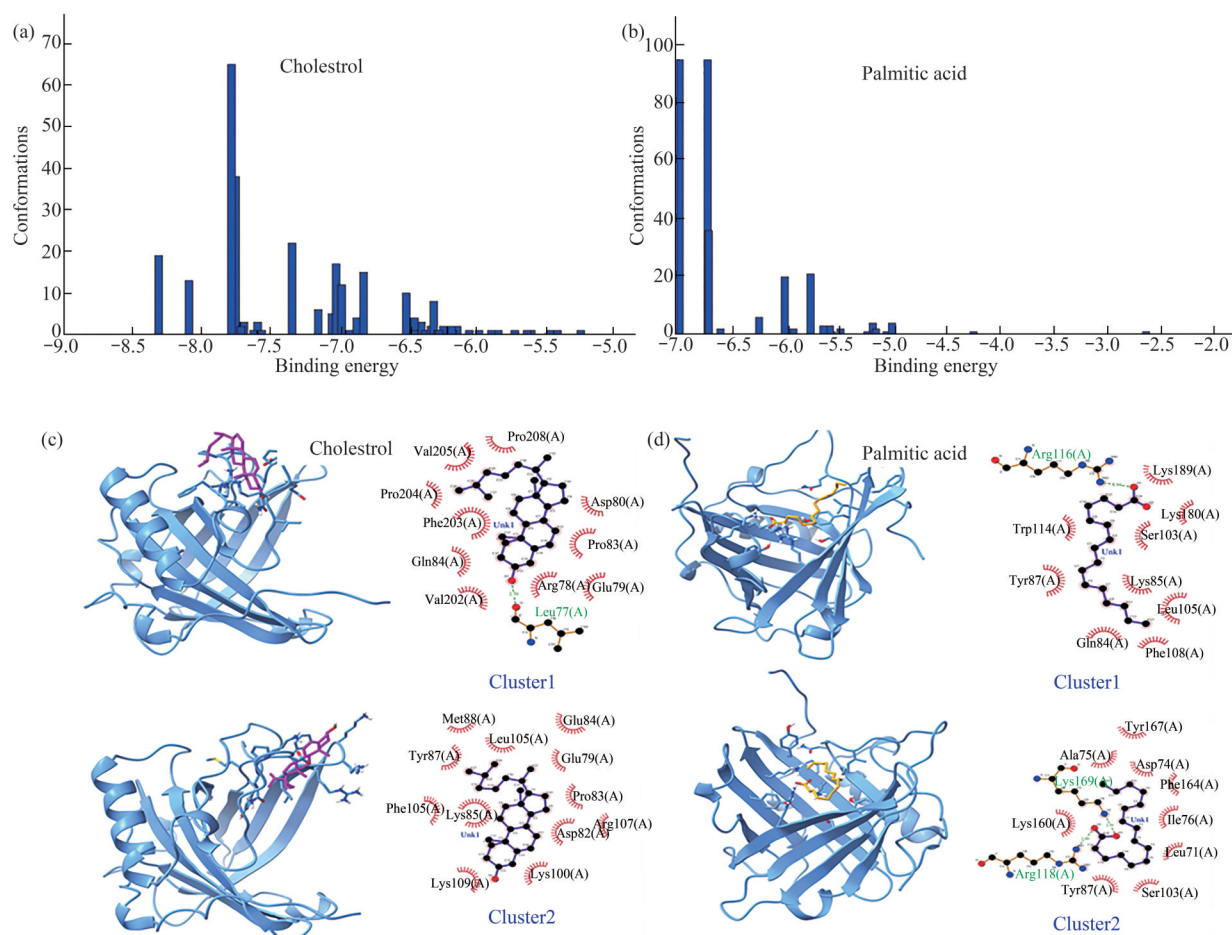


Fig. 5 Binding energies and binding sites analysis of lipocalin–ligand complexes

The number of clusters, the size of each cluster and the amount of best binding energy resulted from molecular docking analysis of lipocalin protein with cholesterol (a) and palmitic acid (b) molecules were analyzed. The binding amino acids involved in the formation of lipocalin-cholesterol (c) and lipocalin-palmitic acid (d) complexes were predicted in clusters 1 and 2 by AutoDock4.2 server.

Table 1 Energy details of the top 3 clusters obtained by molecular docking of cholesterol and palmitic acid with DUF538 (upper) and lipocalin (lower) proteins

Ligand	Cluster	No. of structures	Binding energy	Ligand efficiency	$K_i/(\mu\text{mol}\cdot\text{L}^{-1})$	Intermol energy	VdW energy	Elec. energy
Chol	Cluster1	113	-8.39	-0.3	0.712	-10.18	-10.11	-0.07
	Cluster2	9	-7.9	-0.28	1.63	-9.69	-9.65	-0.04
	Cluster3	52	-7.82	-0.28	1.86	-9.61	-9.54	-0.07
Palm	Cluster1	55	-5.42	-0.3	105.71	-9.6	-7.61	-1.99
	Cluster2	41	-5.29	-0.29	131.76	-9.47	-7.12	-2.35
	Cluster3	2	-5.29	-0.29	132.64	-9.47	-6.82	-2.64
Chol	Cluster1	19	-8.31	-0.3	0.811	-10.1	-10.12	0.02
	Cluster2	13	-8.09	-0.29	1.17	-9.88	-9.82	-0.06
	Cluster3	65	-7.78	-0.28	1.98	-9.57	-9.50	-0.07
Palm	Cluster1	95	-6.99	-0.39	7.57	-11.6	-7.87	-3.29
	Cluster2	95	-6.73	-0.37	11.59	-10.91	-7.54	-3.37

Pro83, Arg78 and Glu79 are involved. In cluster 2, the amino acids Met86, Leu105, Tyr87, Lys85, Phe105, Lys109, Lys100, Asp82, Arg107, Pro83, Glu79 and Glu84 constitute the binding amino acids. In case of lipocalin-palmitic acid complex formation, the amino acids Arg116, Trp114, Tyr87, Gln84, Phe106, Leu105, Lys85, Ser103, Lys160 and Lys169 are involved in cluster 1 in which Arg116 is interacted with hydrogen bonds. In cluster 2, the amino acids Arg116, Lys169, Lys160, Ala75, Tyr167, Asn74, Phe164, Ile76, Leu71, Ser103 and Tyr87 are involved in complex formation in which the amino acids Arg116 and Lys169 are interacted with hydrogen bonds. Data analysis showed that cholesterol and palmitic acid molecules bind to lipocalin protein at its interior parts, as compared to DUF538 protein complexes. Besides this, the comparison of interactive amino acids between four different lipocalin-ligand complexes indicated that the primary structure of test protein has no decisive role in ligand-protein interaction.

Finally, the comparison of DUF and lipocalin ligand-binding abilities indicated that both test proteins are able to bind to cholesterol and palmitic acid molecules with higher binding energy but no significant conserved binding sites. The primary and secondary binding sites were predicted to be varied between DUF lipocalin complexes.

2.3 Experimental analysis of DUF538 binding potential

For further confirmation of the docking results, the binding potential of *Arabidopsis* DUF538 protein towards cholesterol molecule was investigated by MBP-based recombinant technology. To do this, the cDNA of DUF538 was amplified from *Arabidopsis thaliana* leaves (grown under normal laboratory conditions) by RT-PCR using specific primer set as described under materials and methods section. The amplified product was analyzed on agarose gel. The results showed that the size and the partial sequence of the amplified cDNA were in perfect agreement with the expected size and the sequence of the already reported *Arabidopsis* cDNA (data not presented). The amplified cDNA was then restricted with *Eco*RI and *Bam*HI enzymes (the restriction sites were already located on 5' ends of primers) and directionally ligated on pMALc2X expression vector and transformed into TB1 strain of *E. coli*. The MBP-fused DUF538 protein was extracted and purified from TB1

expressing cells by using single-step maltose-affinity chromatography. To confirm the purity of the recombinant product, SDS-PAGE electrophoresis and Western blotting were performed using antibody against maltose-binding protein (Figure 6). The molecular size of the fused protein was analyzed to be about 62 ku on the gel.

In order to assess the cholesterol binding ability of MBP-fused product of DUF538, the free cholesterol content of test solution was assessed before and after mixing with purified fused DUF538 protein and passing through the maltose affinity column. The changes in the absorbance of test samples (before and after mixing with fused protein) was estimated and presented as the approximate binding capacity of fused DUF538 towards cholesterol molecules. The results showed that the absorbance of the test sample is decreased to 0.7 as compared to 1.8, the absorbance of the control samples. Accordingly, the amount of cholesterol was found to be sharply decreased to the level of 39% in comparison to the control (Figure 6). On the other hand, this data revealed that about 61% of the cholesterol molecules are absorbed by DUF538 protein. Comparison of the MBP containing control sample with MBP-free control and test sample

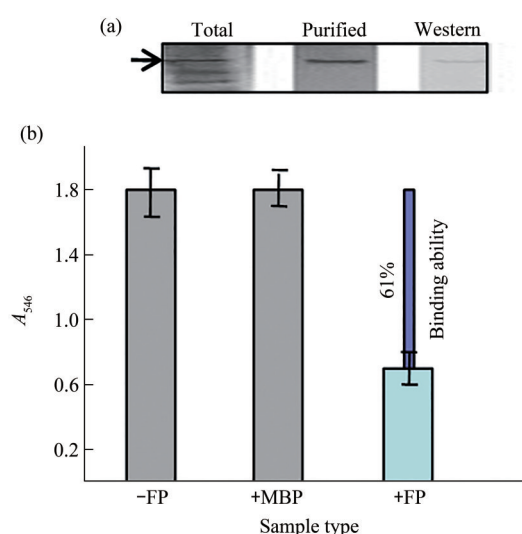


Fig. 6 SDS-PAGE, Western analysis and cholesterol binding ability assessment

(a) SDS-PAGE analysis of total protein extract of recombinant bacteria along with purified fused product of DUF538 and Western analysis. (b) Cholesterol binding ability assessment in the absence of fused product (-FP), in the presence of MBP (+MBP), in the presence of fused product (+FP) showing 61% cholesterol binding ability of DUF538 protein.

showed that cholesterol binding ability is exerted from DUF538 moiety of the fused product and not from the MBP partner. On the other hand, MBP is not involved in cholesterol binding ability of DUF538.

3 Discussion

The aim of this study as the first ever time research work was to analyze the structural similarity as well as the binding potentials between DUF538 and lipocalin proteins using computational as well as experimental tools including recombinant techniques. Based on our previous research results, DUF538 domain containing proteins were suggested to be able to bind to lipid-like molecules such as chlorophyll and lipopolysaccharides^[10-11]. Thereby, DUF538 protein family was proposed to be the structural/functional homologue of lipocalin (lipid binding anticalin like proteins) protein family in plants or animals. Our primary structural analysis did not support the significant structural similarity between DUF538 and plant/animal lipocalin proteins. In accordance with this, the previous studies on sequence and structure relationship been suggested that lipocalins display very low levels of sequence identities often falling below 20%^[13]. In contrast to their low level of conservation at the primary sequences, comparison of different lipocalins tertiary structures have shown that their overall folding pattern including 8 anti parallel β -stranded structure is highly conserved^[17, 21]. Therefore, as a strong clue about the structural/functional similarities between selected test proteins, our searches were extended to identify and compare the tertiary structures of test proteins. The relevant computational analysis, about the similar β -dominated folding pattern, highlighted that DUF538 domain containing proteins might be the potential structural/functional homologues of lipocalins. Our motif analysis data also speculated that there are no extra main motifs on test proteins and DUF538 and lipocalin motifs may be the substitutes of each other. Lipocalin members are typically known as binding proteins. They have different molecular recognition properties including their binding to several lipid-like hydrophobic molecules (such as retinol, fatty acids and cholesterol), binding to specific cell-surface receptors and formation of different macromolecular complexes^[22-23]. Keeping this in view, for more similarity analysis, the binding potentials of DUF538

and lipocalin test proteins were investigated. The overall analysis of molecular docking results regarding the binding potentials of DUF538 and lipocalin proteins indicated that both these proteins are able to bind cholesterol and palmitic acid molecules. However, their abilities to interact with cholesterol molecule are more than palmitic acid. The N-terminus parts of both proteins were found to be the interactive sites for cholesterol and palmitic acid. All these data together revealed that DUF538 protein family can be considered as the potential structural and may be functional homologue of lipocalin superfamily. But to how extend this computational prediction can be reliable, was depended on experimental analysis, too. Therefore, following the computational clarifications we extended our analysis by recombinant-based experimental tool and examined the cholesterol binding ability of DUF538, *in vitro*. The experiment result confirmed the cholesterol binding capacity of DUF538 protein as it has already been reported for lipocalins^[22]. We selected MBP fusion technology since MBP fusion is known as the most efficient production and purification system of correctly folded proteins in heterologous cells, enhancing protein structural folding and stability in a native confirmation^[26-28]. Thereby, the biological activities of MBP-fused products including DUF538 are rarely influenced^[1, 29]. The overall result obtained by this assay did more support our computational data indicating DUF538 family might be the potential functional homologue of lipocalin protein superfamily.

Our previous report suggested that plant DUF538 proteins are the structural/functional homologues of BPI (bactericidal/permeability increasing) proteins in the innate immune system of mammals^[9]. They do affect the bacterial growth rates through the binding to the lipid A moiety of bacterial membranes lipopolysaccharides as like as BPI in mammals immune system. On the other hand, the members of lipocalins are identified as innate immune proteins with bactericidal activity^[30-31]. This reveals that lipocalins and DUF538 are not extra domains, but they can be the potential substitute of each other that share the similar structure and function in different organisms including plants and animals.

The present work results indicated that similar to lipocalins, DUF538 domain containing proteins exhibit efficient binding capacity towards lipid-like

cholesterol molecule. On the other hand, DUF538 proteins were previously found to be the functional homologues of WSCP (water soluble-chlorophyll binding proteins) and efficiently bind to chlorophyll molecules same to WSCP^[11]. Considering chlorophyll molecules containing lipid moiety in their structures, the variable lipid binding and transporting capacities is predicted for DUF538 as been already approved for lipocalins^[22]. Besides binding capacities, DUF538 and lipocalins are also proposed to exert hydrolytic activities towards bound molecules. DUF538 proteins have already been suggested to be structurally and functionally comparable to lipolytic hydrolases particularly esterase-lipases^[10, 12]. This reveals that DUF538 protein family not only is able to bind lipid/lipid-like molecules but also it can use them as substrate for hydrolysis. The hydrolytic activity of DUF538 proteins towards chlorophyll molecules and methylester compounds have already been approved by our research team^[10-12]. Keeping this information in view and considering the structural similarities of DUF538 and lipocalin proteins, their hydrolytic capacities towards cholesterol molecules cannot be away from thinking. We hope the results of the present work will open the gate to specific and interesting studies on cholesterol transporting and metabolizing potencies of DUF538/lipocalin proteins in plant and animal systems. Binding proteins are essential for most cellular events and regulations. To exert their functions, they bind or interact with different other molecular partners that are important in understanding the mechanisms of action in biological world. Thereby, based on the binding abilities of plants DUF538 proteins, they can be nominated as novel binding proteins and may be used in plants, animal or medical biotechnology and drug development.

4 Conclusion

Considering the similar structure and binding potentials, plants DUF538 domain containing protein was proposed to be the structural and may be functional homologues of lipocalin proteins in plants and animals. They were suggested to be the potential substitute of each other and share the similar structure and function in different organisms. The identifications presented in this work will contribute to our future specific investigations and detailed

understanding of the interactions of DUF538 / lipocalin proteins with cholesterol in plant and animal systems.

Acknowledgments The author of this paper is thankful to the Department of Animal Sciences, Faculty of Natural Sciences, and Research Center for Bioscience and Biotechnology, University of Tabriz for the facilities and financial supports.

References

- [1] Gholizadeh A. Heterologous expression of stress-responsive DUF538 domain containing protein and its morpho-biochemical consequences. *Protein J*, 2011, **30**(5): 351-358
- [2] Takahashi S, Yoshikawa M, Kamada A, *et al.* The photoconvertible water-soluble chlorophyll-binding protein of *Chenopodium album* is a member of DUF538, a superfamily that distributes in Embryophyta. *J Plant Physiol*, 2013, **170**(17): 1549-1552
- [3] Araujo A C G, Guimaraes P M, Mota A P Z, *et al.* Overexpression of *DUF538* from wild *Arachis* enhances plant resistance to *Meloidogyne* spp. *Agronomy*, 2021, **11**: 559
- [4] Gholizadeh A, BaghbanKohnehouz B. Identification of DUF538 cDNA clone from *Celosia cristata* expressed sequences of none stressed and stressed leaves. *Russ J Plant Physiol*, 2010, **57**: 247-252
- [5] Li L, Xie C, Ye T, *et al.* Molecular characterization, expression pattern and function analysis of the rice OsDUF866 family. *Biotechnol Biotechnol Equip*, 2017, **31**: 243-249
- [6] Li L H, Lv M M, Li X, *et al.* The rice OsDUF810 family: OsDUF810.7 may be involved in the tolerance to salt and drought. *Mol Biol (Mosk)*, 2018, **52**(4): 489-496
- [7] Nabi R B S, Tayade R, Imran Q M, *et al.* Functional insight of nitric-oxide induced DUF genes in *Arabidopsis thaliana*. *Front Plant Sci*, 2020, **11**: 1041
- [8] Yu C Y, Sharma O, Nguyen P H T, *et al.* A pair of DUF538 domain-containing proteins modulates plant growth and trichome development through the transcriptional regulation of *GLABRA1* in *Arabidopsis thaliana*. *Plant J*, 2021, **108**(4): 992-1004
- [9] Gholizadeh A, Baghbankohnehouz S. DUF538 protein super family is predicted to be the potential homologue of bactericidal/permeability-increasing protein in plant system. *Protein J*, 2013, **32**(3): 163-171
- [10] Gholizadeh A. DUF538 protein superfamily is predicted to be chlorophyll hydrolyzing enzymes in plants. *Physiol Mol Biol Plants*, 2016, **22**(1): 77-85
- [11] Gholizadeh A. Chlorophyll binding abilities of WSCP and DUF538 proteins: the possibility of covalent bonds. *Plant Cell Biotech Mol Biol*, 2017, **18**: 373-380
- [12] Gholizadeh A. Pectin methylesterase activity of plant DUF538 protein superfamily. *Physiol Mol Biol Plants*, 2020, **26**(4):

- 829-839
- [13] Flower D R. The lipocalin protein family: structure and function. *Biochem J*, 1996, **318**(Pt 1): 1-14
- [14] Weiss G A, Lowman H B. Anticalins versus antibodies: made-to-order binding proteins for small molecules. *Chemist Biol*, 2000, **7**(8): R177-184
- [15] Charron J B F, Breton G, Badawi M, *et al.* Molecular and structural analyses of a novel temperature stress-induced lipocalin from wheat and *Arabidopsis*. *FEBS Lett*, 2002, **517**(1-3): 129-132
- [16] Jaber S A, Cohen A, D'Souza C, *et al.* Lipocalin-2: structure, function, distribution and role in metabolic disorders. *Biomed Pharmacother*, 2021, **142**: 112002
- [17] Du Z P, Li B W, Wu X, *et al.* A systematic analysis of human lipocalin family and its expression in esophageal carcinoma. *Sci Rep*, 2015, **5**: 12010
- [18] Zhang X, Cui Y, He M, *et al.* Lipocalin-1 expression as a prognosticator marker of survival in breast cancer patients. *Breast Care (Basel)*, 2019, **15**(3): 272-280
- [19] Charron J B F, Ouellet F, Pelletier M, *et al.* Identification, expression, and evolutionary analyses of plant lipocalins. *Plant Physiol*, 2005, **139**(4): 2017-2028
- [20] Lakshmi B, Mishra M, Srinivasan N, *et al.* Structure-based phylogenetic analysis of the lipocalin family. *PLoS One*, 2015, **10**(8): e0135507
- [21] Flower D R, North A C T, Sansom C E. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta*, 2000, **1482**(1-2): 9-24
- [22] Skerra A. Alternative binding proteins: anticalins-harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS J*, 2008, **275**(11): 2677-2683
- [23] Glasgow B J. Tear lipocalin and lipocalin-interacting membrane receptor. *Front Physiol*, 2021, **12**: 684211
- [24] Bugos R C, Hieber A D, Yamamoto H Y. Xanthophyll cycle enzymes are members of the lipocalin family, the first identified from plants. *J Biol Chem*, 1998, **273**(25): 15321-15324
- [25] Laemmli U K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 1970, **227**(5259): 680-685
- [26] Makrids S C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microb Rev*, 1996, **60**(3): 512-538
- [27] Anderson D C, Krummen L. Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol*, 2002, **13**(2): 117-123
- [28] Zhang L T, Yao Y M, Lu J Q, *et al.* Recombinant bactericidal/permeability-increasing protein inhibits endotoxin-induced high-mobility group box 1 protein gene expression in sepsis. *Shock*, 2008, **29**(2): 278-284
- [29] Kapust B, Waugh D S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 1999, **8**(8): 1668-1674
- [30] Kjeldsen L, Bainton D F, Sengelov H, *et al.* Identification of neutrophil gelatinase-associated lipocalin as a novel matrix protein of specific granules in human neutrophils. *Blood*, 1994, **83**(3): 799-807
- [31] Dahl S L, Woodworth J S, Lerche C J, *et al.* Lipocalin-2 functions as inhibitor of innate resistance to *Mycobacterium tuberculosis*. *Front Immunol*, 2018, **9**: 2717