

# Plant Lipocalins

Jean-Benoit F. Charron and Fathey Sarhan\*

## Summary

Lipocalins are widely distributed in animals, insect and bacteria but very little is known about plant lipocalins. The first lipocalin-like proteins reported in plants were the two key enzymes of the xanthophyll cycle, the violaxanthin de-epoxidases and the zeaxanthin epoxidases. However, the peculiar architecture of these proteins raised doubt as of their true belonging to the lipocalin family. We recently reported the identification and cloning of the first true plant lipocalins from wheat and *Arabidopsis*. The encoded proteins were named temperature-induced lipocalins and possess the three structurally-conserved regions that characterize lipocalins. Sequence analyses revealed that these plant lipocalins share significant homology with three evolutionarily-related lipocalins, the mammalian apolipoprotein D, the bacterial lipocalin Blc and the insect Lazarillo protein. Data mining of genomic databases and bioinformatic predictions revealed that plants possess two other lipocalin members: temperature-induced lipocalin-2 and chloroplastic lipocalin. Expression and regulation studies suggest that the plant lipocalins are associated with environmental stresses.

## Introduction

Lipocalins are an ancient and functionally diverse family of mostly extracellular proteins.<sup>1</sup> This family has been studied in details in bacteria, invertebrates and vertebrates, and these studies have been summarized in several excellent reviews.<sup>2-5</sup> However, very little is known about plant lipocalins.<sup>6-7</sup> The rapidly expanding area of functional, structural and comparative genomics provides opportunities for the identification of lipocalin homologs in plants. Using an integrated approach of data mining of EST databases, bioinformatics predictions, phylogenetic studies, and structural, cellular localization and expression profiling analyses, we identified novel plant lipocalins. Here we describe the molecular characterization and evolution of plant lipocalins and discuss their putative function during plant development under environmental stresses.

## Temperature-Induced Lipocalins

The first true plant lipocalins were recently identified from wheat and *Arabidopsis thaliana*.<sup>7</sup> A full length clone was first isolated from a cDNA library prepared from cold-acclimated wheat tissues and named *TaTIL* for *Triticum aestivum* temperature-induced lipocalin. This gene has since been renamed *TaTIL-1*. The open reading frame encodes a protein of 190 amino acids (aa) with a calculated molecular mass of 22 kDa and a theoretical pI of 5.5 (Table 1). A search in the GenBank ESTs database revealed homology (74% identity, 83% similarity) with a predicted putative protein from *Arabidopsis thaliana* that

\*Corresponding Author: Fathey Sarhan—Université du Québec à Montréal, Département des Sciences Biologiques, C.P. 8888, Succ. Centre-ville, Montréal QC, H3C 3P8, Canada.  
Email: sarhan.fathey@uqam.ca

---

*Lipocalins*, edited by Bo Åkerström, Niels Borregaard, Darren R. Flower and Jean-Philippe Salier. ©2005 Eurekah.com.

---

Table 1. Structural features of plant lipocalins and lipocalin-like proteins

Protein	Precursor/ Mature Molecular Mass (kDa)	Subcellular Localisation	Cleavage Site Position*	SCR1 GxWY	SCR2 TDY	SCR3 R	Conserved Cys Residues	Conserved N-glycosyl. Sites	Other Domains
AfTIL-1	21 / 20	membrane	C-terminal	yes	D only	yes	0	1	no
OsTIL-1	22 / 20	membrane	C-terminal	yes	D only	yes	0	1	no
TaTIL-1	22 / 20	membrane	C-terminal	yes	D only	yes	0	1	no
OsTIL-2	21 / 19	ND	C-terminal	yes	D only	yes	0	1	no
AfCHL	39 / 26	chloroplast	N-terminal	yes	yes	yes	8	0	no
OsCHL	37 / 26	chloroplast	N-terminal	yes	yes	yes	8	0	no
AvDE	52 / 40	chloroplast	N-terminal	yes	no	yes	14	1	yes**
OsVDE	50 / 40	chloroplast	N-terminal	yes	no	yes	14	1	yes**
TaVDE	52 / 40	chloroplast	N-terminal	yes	no	yes	14	0	yes**
AfZEP	74 / 68	chloroplast	N-terminal	yes	no	no	6	1	yes***
OsZEP	68 / 63	chloroplast	N-terminal	yes	no	no	5	1	yes***

At, *Arabidopsis thaliana*; Ta, *Triticum aestivum* (wheat); Os, *Oryza sativa* (rice); Cys, Cysteine; ND, not determined. \* C-terminal, GPI anchor site; N-terminal, signal peptide. \*\* N-terminal cyteine-rich region and C-terminal glutamic acid-rich region. \*\*\* N-terminal ADP-binding site and C-terminal FAD-binding site

we named *At*TIL for *Arabidopsis thaliana* temperature-induced lipocalin. Sequence analysis of this *Arabidopsis* clone revealed that the cDNA encodes a 186 aa protein. The SCR1 region is located from aa 15 to 31 (GLDVARYMGRWYEIASF) in *Ta*TIL-1 and from aa 12 to 28 (GLNVERYMGRWYEIASF) in *At*TIL, and possesses the two conserved amino acids G and W (Table 1).<sup>8-9</sup> The SCR2 of *Ta*TIL-1 is found in the C-terminal portion of the protein from aa 105 to 119 (YWVLYVDDDDYQYALV) while in *At*TIL it is found from aa 101 to 115 (YWVLYIDPDYQHALL). The SCR2 of animal and bacterial lipocalins generally contains a TDY triplet.<sup>8-9</sup> However, in *Ta*TIL-1 and *At*TIL, only the central D is present (Table 1). SCR3 is found in the C-terminal portion of both proteins, from aa 129 to 144 (ILCRKTHIEEEVNQL) in *Ta*TIL-1 and from aa 125 to 140 in *At*TIL (ILSRTAQMEEETYKQL). The conserved R residue that characterizes this fingerprint is present in both sequences (Table 1).<sup>8-9</sup> Further sequence analysis of *Ta*TIL-1 and *At*TIL indicated the presence of a putative N-glycosylation site (Table 1). Putative C-terminal cleavage sites are predicted by several targeting peptide prediction programs (DGPI, PSORT, and SignalP) to be at aa 172 in *Ta*TIL-1 and at aa 168 in *At*TIL.<sup>10-11</sup> Considering this putative cleavage site, the calculated molecular mass of the mature proteins in wheat and *Arabidopsis* is 20 kDa with a pI of 5.2 (Table 1).

The homology search revealed that *Ta*TIL-1 (accession no. AY077702) and its ortholog from *Arabidopsis* (accession no. AY062789) share significant similarity with three evolutionarily related lipocalins: the human apolipoprotein D (ApoD) precursor (accession no. P05090), the *Escherichia coli* outer membrane lipoprotein Blc precursor (accession no. P39281), and the American grasshopper Lazarillo precursor (accession no. P49291). These proteins respectively share 29%, 31%, and 23% identity, and 46%, 54% and 40% similarity with *Ta*TIL-1. Among all lipocalins, Blc, ApoD, and Lazarillo are the only ones known to be anchored to biological membranes.<sup>3</sup> The good similarity between these proteins and the plant TILs suggests that *Ta*TIL-1 and *At*TIL are also membrane-associated proteins. The sequence analysis also revealed that, like the *E. coli* Blc, *Ta*TIL-1 and *At*TIL differ from most lipocalins by the absence of intramolecular disulfide bonds. However, they are potentially N-glycosylated like human ApoD and Lazarillo. When the three SCRs of these five proteins are aligned, the start codons from *Ta*TIL-1 and *At*TIL are positioned at the cleavage sites of the N-terminal signal peptides of the three other proteins. This alignment suggests that *Ta*TIL-1 and *At*TIL do not possess an N-terminal signal peptide as is the case in Blc, ApoD and Lazarillo. The N-terminal portion of *Ta*TIL-1 is composed of hydrophilic residues followed by few hydrophobic residues. In *At*TIL, the hydrophobic section is even less accentuated. This profile does not fit the standard hydrophobic nature of the N-terminal signal peptide identified in ApoD, Blc and Lazarillo. Like Lazarillo, the *Ta*TIL-1 and *At*TIL proteins are longer than ApoD and Blc at their C-terminal end and possess a similar putative cleavage site. The hydrophobic C-terminal tail enables Lazarillo to receive a glycosylphosphatidylinositol (GPI) anchor.<sup>12</sup> This suggests that *Ta*TIL-1 and *At*TIL could also receive a GPI anchor. GPI anchoring is a post-translational addition of a lipid occurring in the endoplasmic reticulum lumen which links proteins to the external face of the plasma membrane. This type of modification has been reported in plants.<sup>13</sup> The fact that the N-glycosylation site is conserved between the wheat and *Arabidopsis* TIL orthologs supports the possibility that these proteins are processed in the endoplasmic reticulum lumen. Another type of attachment to the membrane can also be suggested for *Ta*TIL-1 and *At*TIL. It has been proposed that human ApoD is associated with the external face of the membrane by a hydrophobic loop.<sup>3,14-15</sup> *Ta*TIL-1 and *At*TIL also possess a hydrophobic stretch of seven amino acids that is inserted into a loop between two  $\beta$ -strands. This hydrophobic stretch is in the loop between  $\beta$ -strands 5 and 6 instead of being in the loop between strands 7 and 8, as is the case in the human ApoD (Fig. 1 B1,C2). It is nevertheless possible that this stretch favours the attachment of TILs to the plasma membrane. The loop scaffold in *Ta*TIL-1 and *At*TIL is two amino acids longer than in the human ApoD and there is a proline at positions 32 and 29 respectively. These modifications suggest that the plant TILs have a different binding specificity. A recent proteomic analysis of highly purified plasma membranes

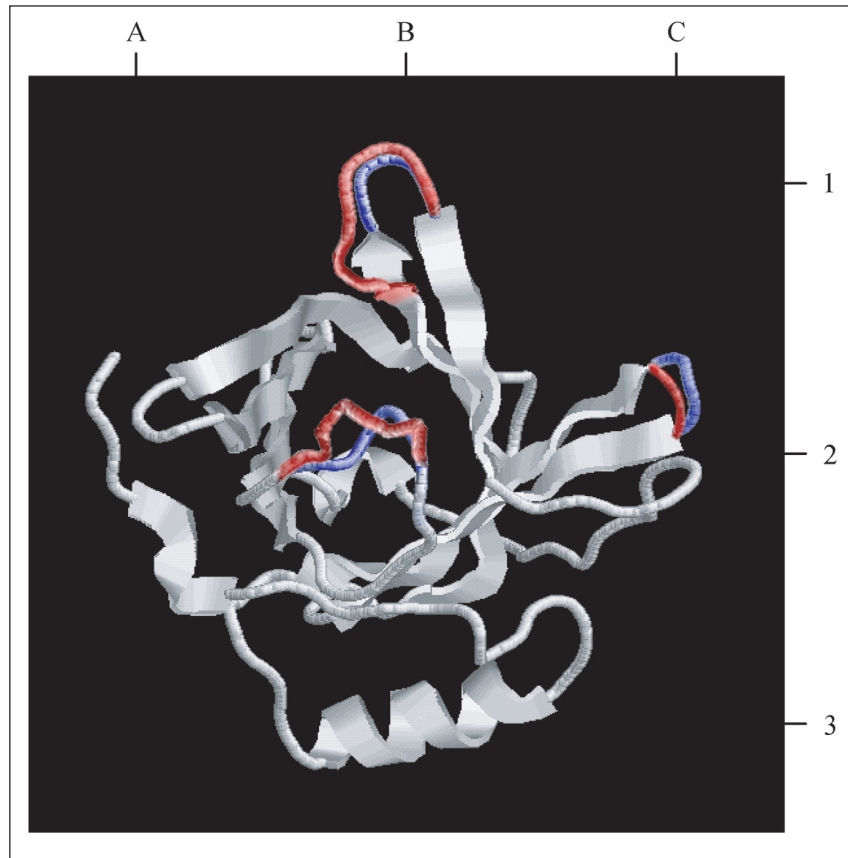


Figure 1. Structural models of human ApoD and wheat *TaTIL-1*. Tertiary structure analyses were carried out using the Swiss-Model program.<sup>16</sup> The lower BLAST limit was set at 0.00001 and the human ApoD model (PDB ID: 2APD)<sup>14</sup> was used as template. The initial result was then resubmitted through the optimizing mode of the program. The final result was then visualized using the Swiss-Pdb Viewer and the model was adapted according to sequence comparison. Differences between the wheat and the human models were superimposed and colored. Grey sections are common to both models. The red (*TaTIL-1*) and blue (ApoD) sections represent structural differences between the two proteins. Reprinted with permission from: Frenette Charron JB, Breton G et al. FEBS Lett 2002; 517(1-3):129-132. ©2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

from *Arabidopsis* showed that *AtTIL* is associated with this membrane fraction.<sup>17</sup> This result confirms the prediction that *TaTIL-1* and *AtTIL* are membrane-associated proteins. However, the nature of the association or attachment is still unknown.

Northern blot analysis revealed that the *TaTIL-1* transcripts accumulate to high levels upon exposure to low temperature and heat-shock treatments (10-fold) and to a lesser extent after a water stress (3.5-fold).<sup>7</sup> Abscisic acid, high salt and wounding treatments have no measurable effect. The *TaTIL-1* transcripts accumulate gradually to a maximum level after 36 days of cold acclimation. Upon deacclimation, the level of transcripts returns to the level seen in the control nonacclimated plants. The accumulation of *TaTIL-1* transcripts in wheat was found to be tissue-specific, as they were detected only in cold-acclimated leaves. The expression analyses revealed that the dicot ortholog *AtTIL* is also induced by low temperature (6-fold) and heat-shock

treatments (9-fold). RNA blot hybridization studies also demonstrated that cold acclimation induces the accumulation of *TaTIL-1* transcripts in both less tolerant and cold hardy wheat. However, this increase is greater in the hardy winter cultivars. Low levels of expression are also found in oat and barley, two less cold tolerant species. This difference in accumulation indicates that the *TaTIL-1* expression is correlated with the plant's capacity to develop freezing tolerance.

Analysis of the promoter regions of *AtTIL*, *TaTIL-1*, *OsTIL-1* and *OsTIL-2* revealed the presence of several low temperature response elements (LTREs), dehydration response elements (DREs) and heat shock elements (HSEs). *TaTIL-1* and *AtTIL* promoter sequences contain more LTREs than the *OsTIL-1* promoter sequence. On the other hand, the *OsTIL-1* promoter contains more HSEs than the *AtTIL* and *TaTIL-1* promoters. This situation is not unexpected since rice does not have the ability to cold acclimate but possesses a higher thermotolerance than wheat and *Arabidopsis*. The fact that TIL promoters possess several light-responsive elements supports the specific expression of the corresponding genes in green photosynthetic leaves.

Temperature stresses are known to induce membrane injuries.<sup>17</sup> The membrane-anchored lipocalins (Blc, ApoD, Lazarillo, and possibly *TaTIL-1* and *AtTIL*) all appear to be expressed in response to conditions that cause membrane stresses, which suggests a biological role in membrane biogenesis and/or repair under severe stress conditions.<sup>5</sup> The plant *TaTIL-1* and *AtTIL* proteins, like the human ApoD, may bind a wide variety of potential ligands of varying structures and functions. The mammalian ApoD is reported to bind arachidonic acid, bilirubin, steroid hormones (progesterone and pregnenolone) and cholesterol.<sup>4</sup> It is interesting to mention that plants also synthesize a wide variety of steroid hormones called brassinosteroids. A treatment with 24-epibrassinolide, a brassinosteroid, increases the tolerance of plants to heat and cold stresses.<sup>18</sup> The enhanced resistance to temperature stress is attributed to increased membrane stability and osmoregulation. It is known that sterol insertion in the plasma membrane increases its fluidity at low temperature and maintains the phospholipids order at high temperature.<sup>19</sup> *TaTIL-1* may be involved in the transport of these sterol molecules to the membrane in response to stress conditions.

### Other Plant Lipocalins

Since plant lipocalins were last reviewed, the sequencing of the *Arabidopsis thaliana* and *Oryza sativa* (rice) genomes has been completed.<sup>6,20-21</sup> The newly identified *TaTIL-1* and *AtTIL* proteins were used to search the proteins predicted from the DNA sequence information of these two genomes using the BLAST program. The search revealed that rice possesses two other lipocalin members, TIL-2 and CHL. Sequence analysis revealed the presence of two different genes in rice encoding TIL lipocalins: *OsTIL-1* and *OsTIL-2* on chromosomes 2 and 8, respectively, whereas *Arabidopsis thaliana* has only *AtTIL* on chromosome 5. The *OsTIL-1* and *OsTIL-2* proteins share 65% identity and 80% similarity. *OsTIL-2* is a protein of 179 aa with a calculated molecular mass of 21 kDa (Table 1). The absence of a N-terminal target peptide suggests that the *OsTIL-2* protein would, like *OsTIL-1*, accumulate in the cytosol. Further sequence analysis of the wheat and rice TIL-2 proteins indicated the presence of a conserved putative N-glycosylation site. In addition, a putative C-terminal cleavage site is predicted by several target peptide prediction programs: DGPI, PSORT,<sup>10</sup> and SignalP.<sup>11</sup> Considering this putative cleavage site, the calculated molecular mass of the mature *OsTIL-2* protein is 19 kDa.

The second new member identified from *Arabidopsis* and rice was named CHL (for chloroplastic lipocalin). This protein was identified in *Arabidopsis* as a putative lipocalin (CAB41869).<sup>6</sup> An homology search revealed that *AtCHL* shares only 23% identity and 40% overall similarity with *AtTIL*. However, a region of 16 amino acids corresponding to SCR1 shows a high similarity with TIL lipocalins. The encoded mature proteins in *Arabidopsis* and rice are respectively 314 aa and 322 aa long with calculated molecular masses of 35 and 36 kDa (Table 1). SignalP

and ChloroP predict N-terminal chloroplastic targeting peptides with high scores in both proteins (Table 1).<sup>11,22</sup> However, the exact length of the chloroplast transit peptide and the location of the proteins within the chloroplast is still unknown. A pairwise sequence alignment predicts chloroplast transit peptide cleavage sites near the beginning of SCR1 in both *At*CHL and *Os*CHL sequences. The mature CHL proteins would have a molecular mass of 26 kDa, which is approximately the usual lipocalin size (Table 1). CHL proteins also possess 8 conserved cysteine residues probably involved in the three-dimensional structure of the protein by forming disulfide bridges. Motif searches against the PROSITE database,<sup>23</sup> after exclusion of patterns with a high probability of occurrence, revealed that *Arabidopsis* and rice CHL proteins possess the SCR1 lipocalin signature (Table 1). This signature perfectly fits the SCR1 consensus used by the ScanProsite software and exhibits the two invariant amino acids G and W that are key features of SCR1.<sup>8-9,24</sup> As in most lipocalins, CHL SCR2 is found in the C-terminal half of the protein and bears the conserved TDY triplet (Table 1).<sup>8-9</sup> SCR3 is also found in the C-terminal portion of both proteins and the conserved R residue that characterizes this fingerprint is present (Table 1).<sup>8-9</sup>

### Violaxanthin De-Epoxidases and Zeaxanthin Epoxidases

Violaxanthin de-epoxidases (VDEs) and zeaxanthin epoxidases (ZEPs) are the most puzzling members with regards to their classification as plant lipocalins. The size and the exon-intron architecture of the genes encoding these enzymes show no significant similarity to the genomic organization of bacterial and animal lipocalin genes and for these reasons, they were not considered as true lipocalins in most studies.<sup>25-26</sup> These enzymes are involved in photoprotection of the photosynthetic apparatus, and are first synthesized as precursor proteins that bear the transit peptide needed for translocation to the thylakoid space of chloroplasts.<sup>6,27</sup> They share the common substrate antheraxanthin and are believed to exhibit similar tertiary structure.<sup>6</sup> VDEs are predicted to be proteins with a central barrel structure flanked by a cysteine-rich N-terminal domain and a glutamate-rich C-terminal domain (Table 1).<sup>28</sup> ZEPs possess ADP-binding and FAD-binding domains and fit the description of a lipocalin based on SCR1 homology (Table 1). Functional analyses of the different domains of VDEs demonstrated that the deletion of any of the cysteine residues in the N-terminal region resulted in a total loss of activity.<sup>28</sup> This is likely because cysteine residues allow the formation of disulfide bridges, which are important determinants of protein conformation. It thus appears that the conformation of the mature protein in the N-terminal portion of VDEs is essential to retain their activities. Deletion analysis of the C-terminal region demonstrated that 71 out of 98 aa could be removed without any loss of activity.<sup>28</sup> However, removal of another 12 aa resulted in a 90% loss of activity and an important reduction of the binding of VDEs to the thylakoid membrane.<sup>28</sup>

Given the feature of VDEs and ZEPs and the strict definition of lipocalins, it is difficult to unequivocally consider these two proteins as true lipocalins. They are at best lipocalin-like proteins that could have arisen from the fusion of an ancestral plant lipocalin to proteins with enzymatic functions.<sup>26,29</sup> Thus, VDEs and ZEPs may represent the first example of lipocalins evolution towards the acquisition of novel functions.

### Evolutionary Origin of Plant Lipocalins and Lipocalin-Like Proteins

To help elucidate the evolutionary origin of plant lipocalins, we investigated the presence of lipocalins and lipocalin-like proteins in algae and cyanobacteria. Algae are considered primitive photosynthetic eukaryotes while cyanobacteria carry a complete set of oxygenic photosynthetic genes. The chloroplast is believed to have evolved from the endosymbiosis of a cyanobacterial ancestor with a eukaryotic host cell. An homology search performed with the *Ta*TIL-1 protein sequence revealed several ESTs from red algae. The search also revealed that cyanobacteria possess a lipocalin gene.

Phylogenetic analyses suggest that TIL lipocalin members were probably inherited from a bacterial gene present in the original host cell, the common ancestor of plants and animals.<sup>1</sup> In some plant species, the TIL-2 lipocalin may have arisen from the duplication of the gene encoding the TIL-1 lipocalin. However, the remaining plant lipocalin and lipocalin-like members CHLs, VDEs and ZEPs might have evolved from a series of duplication of the cyanobacterial ancestor gene after cyanobacteria endosymbiosis from which the chloroplast originated. VDE and ZEP sequences subsequently diverged and acquired new cellular function as xanthophylls cycle enzymes.

## Conclusion

The identification and characterization of plant lipocalins and lipocalin-like proteins will help in designing experiments aimed at the understanding of their cellular function in plants and their role in modulating the responses to temperature and oxidative stresses. Using forward and reverse genetics in the model system *Arabidopsis* should provide the information needed to elucidate the function of each protein in the plant metabolism. In addition, microarray analyses will help in the identification of the target genes associated with over / under expression of the different proteins. The ease with which plants can be manipulated and the availability of mutants are tremendous tools that should enable us to understand the cellular function of lipocalins and lipocalin-like proteins in plants. This information could even help understand the cellular function of lipocalins in mammals.

## Acknowledgements

This work was supported by a Natural Sciences and Engineering Research Council of Canada discovery grant, and by Genome Canada, Genome Québec, and Genome Prairie grants to F. Sarhan. We thank Dr. François Ouellet for helpful discussions and editorial help.

## References

1. Sánchez D, Ganfornina MD, Gutiérrez G et al. Exon-intron structure and evolution of the lipocalin gene family. *Mol Biol Evol* 2003; 20(5):775-783.
2. Akerstrom B, Flower DR, Salier JP. Lipocalins: Unity in diversity. *Biochim Biophys Acta* 2000; 1482(1-2):1-8.
3. Bishop RE. The bacterial lipocalins. *Biochim Biophys Acta* 2000; 1482(1-2):73-83.
4. Rassart É, Bedirian A, Do Carmo S et al. Apolipoprotein D. *Biochim Biophys Acta* 2000; 1482(1-2):185-198.
5. Sánchez D, Ganfornina MD, Bastiani MJ, Lazarillo, a neuronal lipocalin in grasshoppers with a role in axon guidance. *Biochim Biophys Acta* 2000; 1482(1-2):102-109.
6. Hieber AD, Bugos RC, Yamamoto HY. Plant lipocalins: Violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim Biophys Acta* 2000; 1482(1-2):84-91.
7. Frenette Charron JB, 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.
8. Flower DR, North AC, Attwood TK. Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci* 1993; 2(5):753-761.
9. Flower DR. The lipocalin protein family: Structure and function. *Biochem J* 1996; 318(Pt1):1-14.
10. Nakai K, Horton P. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* 1999; 24(1):34-36.
11. Nielsen H, Engelbrecht J, Brunak S et al. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* 1997; 8(5-6):581-599.
12. Ganfornina MD, Sánchez D, Bastiani MJ, Lazarillo, a new GPI-linked surface lipocalin, is restricted to a subset of neurons in the grasshopper embryo. *Development* 1995; 121(1):123-134.
13. Morita N, Nakazato H, Okuyama H et al. Evidence for a glycosylinositolphospholipid-anchored alkaline phosphatase in the aquatic plant *Spirodela oligorrhiza*. *Biochim Biophys Acta* 1996; 1290(1):53-62.
14. Peitsch MC, Boguski MS. Is apolipoprotein D a mammalian bilin-binding protein? *New Biol* 1990; 2(2):197-206.

15. Bishop RE, Penfold SS, Frost LS et al. Stationary phase expression of a novel *Escherichia coli* outer membrane lipoprotein and its relationship with mammalian apolipoprotein D. Implications for the origin of lipocalins. *J Biol Chem* 1995; 270(39):23097-23103.
16. Peitsch MC. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem Soc Trans* 1996; 24(1):274-279.
17. Kawamura Y, Uemura M. Mass spectrometric approach for identifying putative plasma membrane proteins of *Arabidopsis* leaves associated with cold acclimation. *Plant J* 2003; 36(2):141-154.
18. Clouse SD, Sasse JM. Brassinosteroids: Essential regulators of plant growth and development. *Annu Rev Plant Physiol Plant Mol Biol* 1998; 49:427-451.
19. Demel RA, De Kruyff B. The function of sterols in membranes. *Biochim Biophys Acta* 1976; 457(2):109-132.
20. The Arabidopsis Initiative 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000; 408:796-815.
21. Yu J, Hu S, Wang J et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 2002; 296(5565):79-92.
22. Emanuelsson O, Nielsen H, von Heijne G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* 1999; 8(5):978-984.
23. Sigrist CJ, Cerutti L, Hulo N et al. PROSITE: A documented database using patterns and profiles as motif descriptors. *Brief Bioinform* 2002; 3(3):265-274.
24. Gattiker A, Gasteiger E, Bairoch A. ScanProsite: A reference implementation of a PROSITE scanning tool. *Appl Bioinformatics* 2002; 1(2):107-108.
25. Gutiérrez G, Ganfornina MD, Sánchez D. Evolution of the lipocalin family as inferred from a protein sequence phylogeny. *Biochim Biophys Acta* 2000; 1482(1-2):35-45.
26. Salier JP. Chromosomal location, exon/intron organization and evolution of lipocalin genes. *Biochim Biophys Acta* 2000; 1482(1-2):25-34.
27. Bugos RC, Hieber AD, Yamamoto HY. Xanthophyll cycle enzymes are members of the lipocalin family, the first identified from plants. *J Biol Chem* 1998; 273(25):15321-15324.
28. Hieber AD, Bugos RC, Verhoeven AS et al. Overexpression of violaxanthin de-epoxidase: Properties of C-terminal deletions on activity and pH-dependent lipid binding. *Planta* 2000; 214(3):476-483.
29. Ganfornina MD, Gutiérrez G, Bastiani M et al. A phylogenetic analysis of the lipocalin protein family. *Mol Biol Evol* 2000; 17(1):114-126.