ISSN: 1314-6246 Filiz & Koç *J. BioSci. Biotech.* **2014**, 3(1): 61-67.

RESEARCH ARTICLE

Ertuğrul Filiz [1]
Ibrahim Koç [2]

In silico sequence analysis and homology modeling of predicted beta-amylase 7-like protein in *Brachypodium distachyon* L.

Authors' addresses:
[1] Department of Crop and Animal Production, Çilimli Vocational School, Düzce University, Çilimli, Düzce 81750, Turkey.
[2] Gebze Institute of Technology, Faculty of Science, Department of Molecular Biology and Genetics, Gebze, Kocaeli 41400, Turkey.

Correspondence:
Ertuğrul Filiz
Department of Crop and Animal Production, Çilimli Vocational School, Düzce University, Çilimli, Düzce 81750, Turkey.
Tel.: +90 380 681 7312 ext. 7413
e-mail: ertugrulfiliz@gmail.com

Article info:
Received: 14 September 2013
Accepted: 29 November 2013

ABSTRACT

Beta-amylase (β-amylase, EC 3.2.1.2) is an enzyme that catalyses hydrolysis of glucosidic bonds in polysaccharides. In this study, we analyzed protein sequence of predicted beta-amylase 7-like protein in *Brachypodium distachyon*. *pI* (isoelectric point) value was found as 5.23 in acidic character, while the instability index (II) was found as 50.28 with accepted unstable protein. The prediction of subcellular localization was revealed that the protein may reside in chloroplast by using CELLO v.2.5. The 3D structure of protein was performed using comparative homology modeling with SWISS-MODEL. The accuracy of the predicted 3D structure was checked using Ramachandran plot analysis showed that 95.4% in favored region. The results of our study contribute to understanding of β-amylase protein structure in grass species and will be scientific base for 3D modeling of beta-amylase proteins in further studies.

Key words: *Brachypodium distachyon*, β-amylase, homology modeling, 3D structure, Swiss-Model

Introduction

Amylase is a kind of enzyme, including catalysis of breakdown of starch into sugars. This enzyme is observed in plants and some bacteria. All types of amylases belong to glycoside hydrolases with related to α-1,4-glycosidic bonds in polysaccharides, including amylose, amylopectin, glycogen, or their degradation products (Dunn, 1974; Oyefuga et al., 2011). One of the major amylases enzyme is β-amylase (α-1,4-glucan maltohydrolase) that catalyzes the liberation of β-anomeric maltose from the non-reducing ends of starch and glycogen (Hirata et al., 2004). Although, distinct conserved regions were identified in several sources, plant and bacterial β-amylase indicates nearly 30% sequence identity.

Many plant β-amylases have been sequenced, such as soybean, barley, rye, *Arabidopsis thaliana*, and sweet potato (Kang et al., 2003). When compared, the soybean and sweet potato β-amylases displayed 67% amino acid sequence identity In addition, the β-amylases of soybean and sweet potato of quaternary structure were notably different (Yoshida & Nakamura, 1991) and they contain different amino acid residues between the β-amylase of soybean and sweet potato (Kang et al., 2003).

Glu^{186} and Glu^{380} residues play important role in the enzymatic reaction in soybean β-amylase (Mikami et al., 1993). Also, the β-amylase has been identified in alfalfa, some forage legumes, pea, and some Solanaceae and Brassicaceae species (van Damme et al., 2001). The β-amylase displays a complicated gene regulation and expression; β-amylase genes are regulated by light, sugars, phytohormones, and abiotic stresses, including salt, cold, and heat stress (Kaplan & Guy, 2004). The *Arabidopsis* includes 9 genes for β-amylase (*BMY*1 to *BMY*9) and only 3 genes (*BMY*7, *BMY*8, and *BMY*9) were accepted coding proteins for chloroplast localization (Monroe & Preiss, 1990).

The aim of this study was to generate predicted 3D structure of β-amylase 7-like protein by using comparative homology modeling. Also, primary and secondary structure analyses were performed with various bioinformatics tools.

RESEARCH ARTICLE

Materials and Methods

Protein sequence data and analysis

The protein sequence of beta-amylase 7-like (accession no: XP_003576871) in *Brachypodium distachyon* was downloaded from NCBI protein database (http://www.ncbi.nlm.nih.gov/protein/). The physicochemical analysis were calculated by ProtParam tool (http://web.expasy.org/protparam/), including *pI*, total number of negatively and positively charged residues, the instability index (II), aliphatic index, and grand average of hydropathicity (GRAVY).

Structural and functional characterization

Secondary structure prediction was performed by using SOPMA (Geourjon & Deléage, 1995) server (http://npsa-pbil.ibcp.fr/). Subcellular localization was predicted by using CELLO v.2.5 (Yu et al., 2004; Yu et al., 2006) 1.1 server (http://www.cbs.dtu.dk). Motif Scan (Pagni et al., 2007; Sigrist et al., 2010) server (http://myhits.isb-sib.ch/cgi-bin/motif_scan) was used to identify known motifs in the sequence. Furthermore, Pfam server (http://www.sanger.ac.uk/software /pfam/search.html) was used for domain analysis (Punta et al., 2012).

Homology modeling and model evaluation

Homology modeling was used for determining 3D structure of protein. Then, BLASTP was performed against PDB (Protein Databank, Bernstein et al., 1977) to retrieve the best suitable templates for homology modeling. PDB ID 2XFR having 52.5% identities were preferred containing maximum identity and lowest e-value that it was used as a template. The modeling of the 3D structure of the protein was performed by using Swiss-Modeler (http://swiss-model.expasy.org/) program (Arnold et al., 2006; Bordoli et al., 2009). After modeling, the quality and validation of the model was evaluated by several structure assessment methods, containing Z-Score by using QMEAN (Benkert et al., 2011), Rampage Ramachandran plot analysis (http://mordred.bioc.cam.ac.uk), and ERRAT (Colovos & Yeates, 1993).

Results and Discussions

The physicochemical analysis of the predicted β-amylase 7-like was performed using Protparam and results were shown in Table 1. This protein had 690 amino acids with molecular weight of 76325.7 Daltons and *pI* of 5.23. The optimum pH of higher plant β-amylases is nearly 5.4, whereas bacterial β-amylases are about 6.7 (Hirata et al., 2004). This pH value (5.4) is so similar to our pH value as 5.23. The most abundant amino acid was found as alanin (78 residues, 11.30%), whereas the lowest was cysteine (12 residues, 1.74%). The total number of negatively charged residues (Asp + Glu, 98) was found higher than the total number of positively charged residues (Arg + Lys, 76). Intracellular proteins have lower number of cysteine residues, but also higher numbers of aliphatic and charged amino acid residues (Nakashima & Nishikawa, 1994). This data is in agreement with our finding that the highest number of amino acid residue was alanine, while the lowest one was cysteine.

It is accepted that extracellular proteins include a more disulphide bridges and cysteine residues (Bradshaw, 1989). The intracellular proteins contain more negatively charged residues (Cedano et al., 1997) and it can be suggested that β-amylase 7-like protein was intracellular portion. Negative GRAVY value shows that this protein is accepted as hydrophilic character. The instability (II) and aliphatic index revealed that this protein may be unstable and globular protein. The predicting of subcellular localization of unknown proteins contributes to understanding of their functions (Idrees et al., 2012), it was performed using CELLO v.2.5 and our protein was localized in chloroplast. β-amylases are observed in the stroma of chloroplasts of the mesophyll cells (Scheidig et al., 2002) and this data is consistent with our findings.

The secondary structure of the protein was predicted using SOPMA server (Table 2). It was observed that random coil was predominant (44.49%), followed by alpha helix (34.06%) and extended strand (14.78%). Also, beta turn was found as 6.67%. Random coils have important functions in proteins for flexibility and conformational changes such as enzymatic turnover (Buxbaum, 2007). Our findings could be related with the enzymatic function of protein.

The domain analysis was conducted using Pfam database and glycosyl hydrolase family 14 was detected. Glycoside hydrolases are commonly known group of enzymes that hydrolyse the glycosidic bond between carbohydrates with more than 100 different families (Henrissat & Davies, 1995). This data support that our protein may play role as enzyme in hydrolysis reactions. The Motif scan tool was used to determine different motifs (Table 3).

RESEARCH ARTICLE

Table 1. *The physicochemical properties of the predicted β-amylase 7-like protein*

Parameters	Value	Explanation
pI	5.23	The protein is accepted as acidic
Total number of negatively charged residues (Asp + Glu) and total number of positively charged residues (Arg + Lys)	98.76	Total numbers of negatively charged residues are higher than the total number of positively charged residues. This protein has intracellular portion.
The instability index (II)	50.28	The protein may be unstable
Aliphatic index	76.54	shows that this globular protein is thermostable
Grand average of hydropathicity (GRAVY)	-0.370	A negative GRAVY score reveals that the protein is hydrophilic

Table 2. *Secondary structure of β-amylase 7-like protein using SOPMA*

Parameters	Number of amino acids	Amino acids (%)
Alpha helix (Hh)	235	34.06
3_{10} helix (Gg)	0	0.00
Pi helix (Ii)	0	0.00
Beta bridge (Bb)	0	0.00
Extended strand (Ee)	102	14.78
Beta turn (Tt)	46	6.67
Bend region (Ss)	0	0.00
Random coil (Cc)	307	44.49
Ambigous states	0	0.00
Other states	0	0.00

Table 3. *The motifs of predicted β-amylase 7-like protein by Motif Scan*

Motif information	No. of sites	Amino acid residues
N-glycosylation site	6	84-87, 265-268, 309-312, 485-488, 580-583, 635-638
cAMP- and cGMP-dependent protein kinase phosphorylation site	1	370-373
Casein kinase II phosphorylation site	6	332-335, 358-361, 365-368, 390-393, 416-419, 582-585
N-myristoylation site	8	215-220, 226-231, 298-303, 337-342, 421-426, 529-534, 576-581, 617-622
Protein kinase C phosphorylation site	6	72-74, 86-88, 123-125, 165-167, 365-367, 401-403
Beta-amylase active site 1	1	334-342
Glycosyl hydrolase family 14	1	256-676

RESEARCH ARTICLE

The seven type's motifs were observed and the highest number of motif was N-myristoylation site with 8 times. N-glycosylation site, cAMP- and cGMP-dependent protein kinase phosphorylation site, casein kinase II phosphorylation site, N-myristoylation site, protein kinase C phosphorylation site, beta-amylase active site 1, and glycosyl hydrolase family 14 were identified as 6, 1, 6, 8, 6, 1, and 1times, respectively. The phosphorylation of a protein can affect functions and activities of proteins, including intrinsic biological activity, half-life, subcellular location, and docking with other proteins (Cohen, 2000). The occurrence of many phosphorylation sites in β-amylase 7-like protein support that it may be regulated frequently. Myristoylation is post-translational protein modification observed in plants, animals, fungi, and viruses; it is performed by attached myristic acid in proteins. Myristoylation can affect conformational stability of proteins by interaction with membranes or the hydrophobic domains of other proteins (Podell & Gribskov, 2004; Zheng et al., 1993; Olsen & Kaarsholm, 2000). Also, myristoylation was identified in many cellular pathways as playing important roles such as signal transduction, apoptosis, and extracellular export of proteins (Podell & Gribskov, 2004). The eight N-myristoylation sites in our proteins showed that it could be arranged conformational stability for various catalytic activities. In addition, beta-amylase active site 1 and glycosyl hydrolase family 14 prove the catalytic activities of our protein.

Protein 3D structure contributes to understanding of protein function and active sites, and facilitating drug design. X -ray crystallography or NMR spectroscopy are difficult and costly process than computational methods (Kopp & Schwede, 2004; Jaroszewski, 2009). The SWISS-MODEL homology modeling program was used for the predicting of three dimensional structure of the β-amylase 7-like protein (Figure 1). PDB 2XFR was selected as template with 52.49% sequence identity to query sequence (XP_003576871).

After model building, the structure was validated through energy minimization with Z-Score by using Qmean server, ERRAT, and Rampage Ramachandran plot analysis. The Z-score is used to estimate the quality of model using structured solved proteins as references (Benkert et al., 2009). Z-Score was found as -0.92 (Figure 2). ERRAT is a protein structure verification algorithm that analyzes statistics of non-bonded interactions between different atom types based on characteristic atomic interaction (Colovos & Yeates, 1993). The overall quality factor was found as 90.61 which is very satisfactory (Figure 3).

The stereochemical quality of the modelled protein was analyzed by RAMPAGE (Figure 4). Ramachandran plot analysis showed that only 1.4% residues in outlier region, 3.2% allowed region and 95.4% in favored region, indicating that the models were of reliable and good quality.

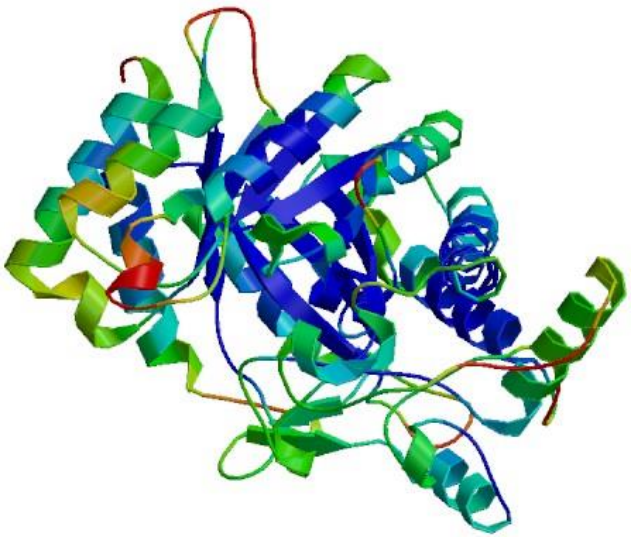

Figure 1. *The three-dimensional structure of predicted beta-amylase 7-like protein of B. distachyon by modelled SWISS-MODEL using PDB ID: 2XFR as template and accession no: XP_003576871 as target*

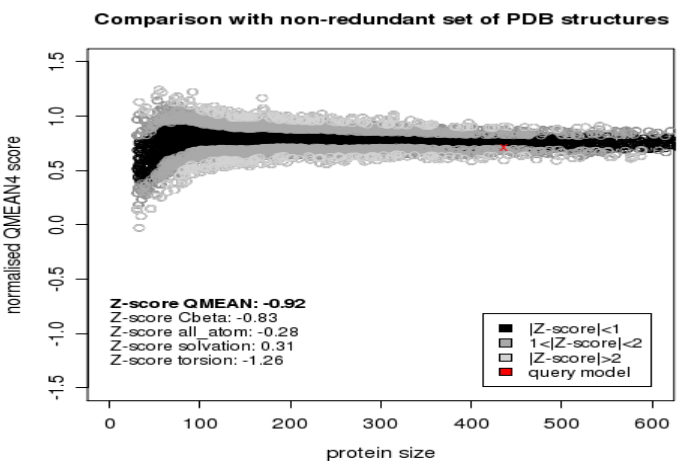

Figure 2. *Z-score of query protein using QMEAN server*

RESEARCH ARTICLE

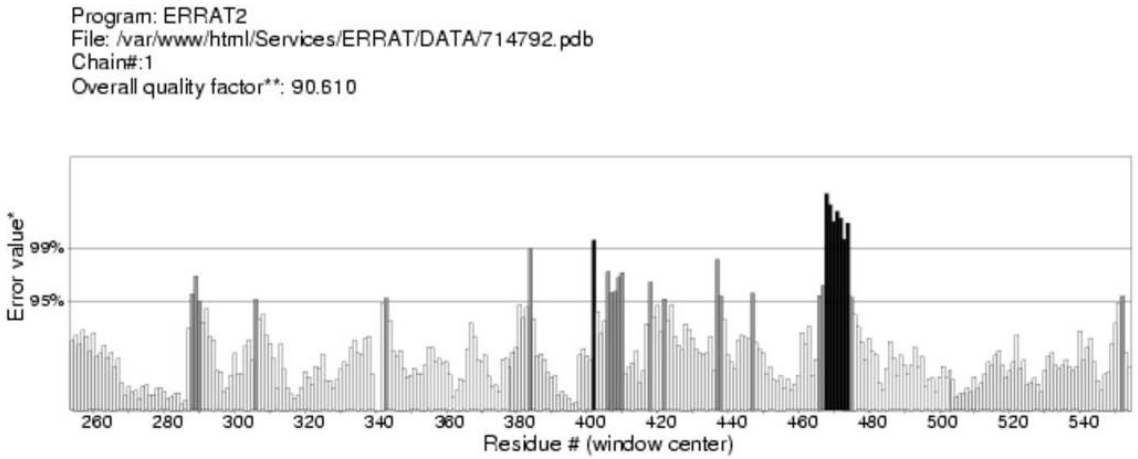

Figure 3. *Overall quality factor evaluated by ERRAT*

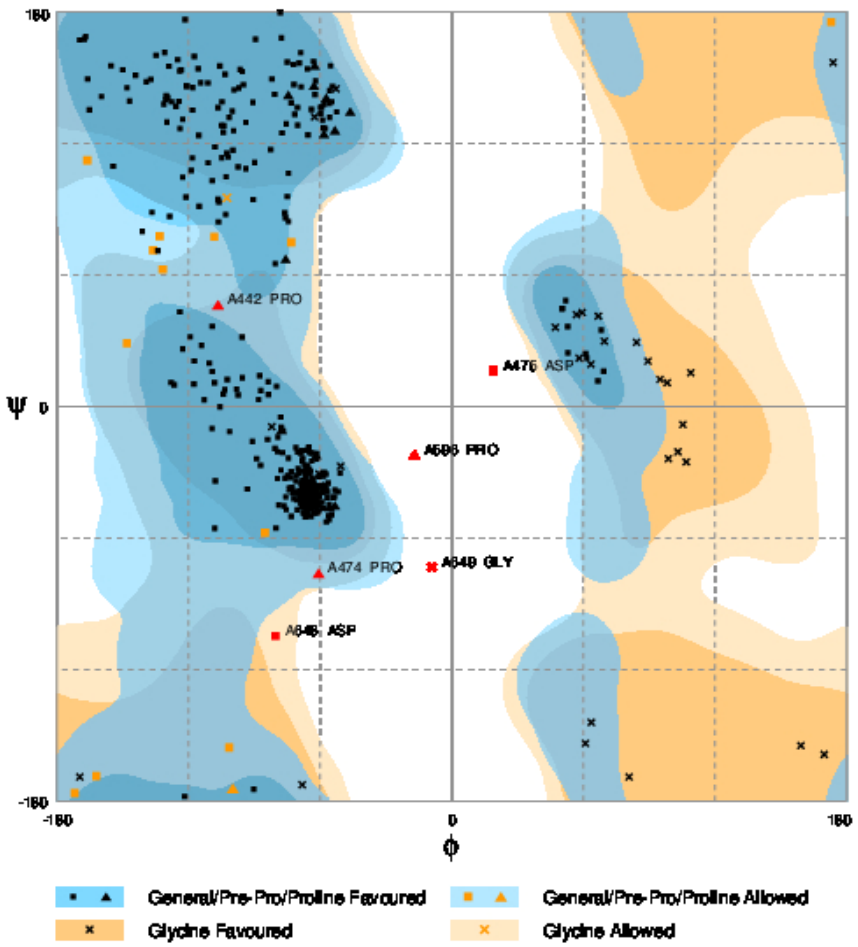

Figure 4. *RAMPAGE values for indicating number of residues in favoured, allowed, and outlier region.*

RESEARCH ARTICLE

Conclusion

The beta-amylase is widely known enzyme to catalyze carbohydrates. In this study, the 3D model of predicted β-amylase 7-like protein was generated by using homology modeling with SWISS-MODEL. The final refined model was further evaluated by using ERRAT, RAMPAGE, and Z-score. The predicted 3D structure will support to understanding of structure of β-amylase proteins in grass species.

References

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics, 2: 195–201.

Benkert P, Biasini M, Schwede T. 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27(3): 343-350.

Benkert P, Künzli M, Schwede T. 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res., 37(Web Server issue): W510-W514.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol., 112: 535–542.

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. 2009. Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols, 1: 1–13.

Bradshaw, RA 1989. Protein translocation and turn-over in eukaryotic cells. Trends Biochem. Sci., 14: 276-279.

Buxbaum, E. 2007. Fundamentals of protein structure and function. -Springer Science Business Media, LLC, New York.

Cedano J, Aloy P, Pérez-Pons JA, Querol E. 1997. Relation between amino acid composition and cellular location of proteins. J. Mol. Biol. 266(3): 594-600

Cohen P. 2000. The regulation of protein function by multisite phosphorylation - a 25 year update. Trends Biochem Sci,. 25: 596-601.

Colovos VC, Yeates TO. 1993. Verification of protein structures: Patterns of non-bonded atomic interactions. Protein Sci, 2: 1511–1519.

Dunn G. 1974. A model for starch breakdown in higher plants, Phytochem., 13: 1341-1346.

Geourjon C, Deléage G. 1995. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci., 11: 681-684.

Henrissat B, Davies G. 1995. Structures and mechanisms of glycosyl hydrolases. Structure, 3 (9): 853–859.

Hirata A, Adachi M, Sekine A, Kang YN, Utsumi S, Mikami B. 2004. Structural and enzymatic analysis of soybean β-amylase mutants with increased pH optimum. The Journal of Biological Chemistry, 279: 7287–7295.

Idrees S, Nadeem S, Kanwal S, Ehsan B, Yousaf A, Nadeem S, Rajoka MI. 2012. *In silico* sequence analysis, homology modeling and function annotation of *Ocimum basilicum* hypothetical protein G1CT28_OCIBA. Int. J. Bioautomation, 16(2): 111-118.

Jaroszewski L. 2009. Protein structure prediction based on sequence similarity. Methods Mol Biol., 569: 129-156.

Kang YN, Adachi M, Mikami B, Utsumi S. 2003. Change in the crystal packing of soybean β-amylase mutants substituted at a few surface amino acid residues. Protein Engineering, 16: 809-817.

Kaplan F, Guy CL. 2004. β-amylase induction and the protective role of maltose during temperature shock. Plant Physiology, 135: 1674–1684.

Kopp J., Schwede T. 2004. Automated protein structure homology modeling: a progress report. Pharmacogenomics, 5(4): 405-416.

Mikami B, Hehre JA, Sato M, Katsube Y, Hirose M, Morita Y, Sacchattini JS. 1993. The 2.0-ANG resolution structure of soybean beta-amylase complexed with alpha-cyclodextrin. Biochemistry, 32: 6836–6845.

Monroe JD, Preiss J. 1990. Purification of a b-amylase that accumulates in *Arabidopsis thaliana* mutants defective in starch metabolism. Plant Physiol, 94: 1033–1039.

Nakashima H, Nishikawa K. 1994. Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. J. Mol. Biol., 238: 54-61.

Olsen HB, Kaarsholm NC. 2000. Structural effects of protein lipida-tion as revealed by LysB29 -myristoyl, des(B30) insulin. Biochemistry, 39: 11893-11900.

Oyefuga OH, Adeyanju MM, Adebawo OO, Agboola FK. 2011. Purification and some properties of β-amylase from the nodes of sugar cane, *Saccharium offinacium*. International Journal of Plant Physiology and Biochemistry, 3(7): 117-124.

Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, Martin O, Kuznetsov D, Falquet L. 2007. MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res., 35: 433 - 437.

Podell S, Gribskov M. 2004. Predicting N-terminal myristoylation sites in plant proteins. BMC Genomics, 5: 37.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. Nucleic Acids Res., Database Issue, 40: D290-D301.

Scheidig A, Fro ¨hlich A, Schulze S, Lloyd JR, Kossmann J. 2002. Down-regulation of a chloroplast-targeted beta-amylase leads to starch-excess phenotype in leaves. Plant J, 30: 581–591

Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. 2010. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res., 38: 161–166.

van Damme EJM, Hu J, Barre A, Hause B, Baggerman G, Rouge P, Peumans WJ. 2001. Purification, characterization, immunolocalization and structural analysis of the abundant cytoplasmic b-amylase from *Calystegia sepium* (hedge bindweed) rhizomes. Eur. J. Biochem., 268: 6263–6273.

Yoshida N, Nakamura K. 1991. Molecular cloning and expression in *Escherichia coli* of cDNA encoding the subunit of sweet potato β-amylase. J. Biochem. (Tokyo), 110: 196-201.

RESEARCH ARTICLE

Yu CS, Chen YC, Lu CH, Hwang K. 2004. Prediction of protein subcellular localization, proteins: Structure, function and bioinformatics, 64: 643-651.

Yu CS, Lin CJ, Hwang JK. 2006. Predicting subcellular localization of proteins for Gram-negative bacteria by Support Vector Machines based on n-peptide compositions. Protein Science, 13: 1402-1406.

Zheng J, Knighton DR, Xuong NH, Taylor SS, Sowadski JM, Ten Eyck LF. 1993. Crystal structures of the myristoylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. Protein Sci,, 2: 1559-1573.