BIOINFORMATICS ORIGINAL PAPER Vol. 27 no. 16 2011, pages 2201–2208
doi:10.1093/bioinformatics/btr371

Structural bioinformatics

Advance access publication: June 27, 2011

In silico analysis of phytoene synthase and its promoter reveals hints for regulation mechanisms of carotenogenesis in *Duanliella bardawil*

Yong-Min Lao[1], Lan Xiao[1], Zhi-Wei Ye[1], Jian-Guo Jiang[1,*] and Shi-Shui Zhou[2]

[1]College of Food and Bioengineering, South China University of Technology, Guangzhou, 510640 and [2]School of Biological Science & Engineering, South China University of Technology, Guangzhou, 510006, China

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Previous researches showed that phytoene synthase (Psy) from *Dunaliella bardawil* is the first regulatory point in carotenogenesis. We hypothesize certain interactions between the environmental stress factors and the regulatory sequences of Psy in *D.bardawil* (DbPsy). Consequently, LA PCR-based genomic walking approach was performed for isolation of *psy* promoter and terminator, respectively. The obtained nucleic acid sequences and the corresponding protein structure of DbPsy were analyzed and predicted using various bioinformatics tools. Finally, we presented some hints for the regulation mechanisms of DbPsy at the molecular level according to the computed results.

Results: LA PCR-based genomic walking results showed that the isolated sequences are the promoter and terminator of *psy*, correspondingly. Computational analysis demonstrated several candidate motifs of the promoter exhibiting hypothetic UV-B-, norglurzon- and salt-induced characteristics, as well as some typical domains universally discovered in promoter sequences, such as TATA-box, CCAAT-box and GATA-box, etc. Furthermore, the structure of Psy was also predicted and aligned along with many counterparts at the protein level. Low homology of N-terminus was found in *D.bardawil*, while a relatively conserved C-terminus was predicted to be involved in the catalytic activity and substrate recognization/binding. Phylogenic analysis classified the DbPsy into a cluster with other algae. These results implied that Psy may share similar regulation mechanisms among algae with respect to their C-termini; while the diversity in N-terminus among Psys, along with the predicted inducible motifs in *psy* promoter from *D.bardawil*, may confer the fine tuning differences between *D.bardawil* and other algae.

Conclusion: By means of computer techniques, we found in *D.barawali* that two interesting conserved motifs of *psy* promoter may involve in UV-B, norglurzon and salt regulation correspondingly; and that the diversity of Psy protein mainly lies in the N-termini among algae. These results indicate some hints for regulation mechanisms of carotenogenesis in *D.bradawil*.

Contact: jgjiang@scut.edu.cn

Received on April 12, 2011; revised on June 2, 2011; accepted on June 12, 2011

*To whom correspondence should be addressed.

1 INTRODUCTION

Carotenoids are a wide family of polyene pigments synthesized by all plants, various algae and certain species of bacteria, fungi and archaea (Goodwin, 1980). In photosynthetic organisms, carotenoids carry out several essential functions, i.e. participating in photosynthesis to protect the photosynthetic apparatus from potential oxidative damage besides light harvesting (Demmig-Adams and Adams, 1996; Siefermann-Harms, 1987). Animals cannot synthesize carotenoids but may obtain them through derivative pigmentation from diet (Olson, 1994). Some carotenoids have vitamin A activity in humans; therefore, they are classified as essential nutrients (Farré *et al.*, 2010). Many researches show the prevention and treatment properties of carotenoids against several kinds of diseases, such as certain chronic disease (Landrum and Bone, 2001), cardiovascular disease (Shaish *et al.*, 2006) as well as certain cancers (Michaud *et al.*, 2000). These promoting properties of carotenoid have attracted intensive interesting in carotenoid biosynthesis and its commercial exploitation. So far, at least 700 carotenoids have been characterized from nature carotenoid biosynthetic pathways involving in C30 and C40 isoprenoids (Feltl *et al.*, 2005).

In recent years, *Dunaliella bardawil* has received intensive attention as a natural rich source of β-carotene, owing to its capability to accumulate massive amount of β-carotene under stress conditions such as high light density, high salt concentration and nutrient starvation (Ben-Amotz *et al.*, 1982). Many studies on accumulation of β-carotene and phytoene have been conducted to date (Ben-Amotz *et al.*, 1987; León *et al.*, 2005; Mogedas *et al.*, 2009; Muthukannan *et al.*, 2010). For large production of β-carotene, the optimal conditions should be high UV-A intensity, slightly alkaline pH, low NaCl concentration and moderate $NaNO_3$ concentration (Mogedas *et al.*, 2009; Muthukannan *et al.*, 2010); while UV-B and other photobleaching compounds could elevate phytoene production, instead of β-carotene (Ben-Amotz *et al.*, 1987; León *et al.*, 2005; White and Jahnke, 2002). Nevertheless, these investigations were primarily focused on culture condition optimization at physiologic and biochemical level, none of them probed into the molecular mechanisms for carotenogenic regulation in *D.bardawil*.

The hypothetic biosynthetic pathway of carotenoid in *Dunaliella* is disproportional to synthesize various carotenoids such as β-carotene, neoxanthin, lutein and so on (Ye *et al.*, 2008). The most universal carotenoid biosynthetic pathway is the successional

reactions leading to the formation of β-carotene. The first committed step is a head-to-head condensation of two geranylgeranyl diphosphate (GGPP) molecules to produce phytoene by phytoene synthase (Psy) (Salvini *et al.*, 2005). Subsequently, phytoene is converted to lycopene through four desaturation reactions, and β-carotene is finally formed by two cyclization steps. Enzymes involved are membrane-associated or integrated into membranes, and Psy is considered to be the first rate-determining point to control the carbon resource flux toward carotenoid synthesis (Shewmarker *et al.*, 1999). At present, many Psy cDNAs have been isolated to investigate the regulation mechanisms controlling carotenoid biosynthesis in plants and microalgae (Salvini *et al.*, 2005; Yan *et al.*, 2005); however, most of them remain confined to molecular cloning and expression analysis. In *D.bardawil*, Psy is even merely cloned and sequenced (GenBank: EU328287.1).

We hypothesize certain interactions between the environmental stress factors and the regulatory sequences of Psy in *D.bardawil* (DbPsy), more specifically, the conserved motifs encompassing the *psy* gene and the resulting diversity of protein structure caused by the variation of amino acid sequence in comparison to other species. Therefore, we isolated and cloned the promoter and terminator of *psy*, and then analyzed the obtained nucleic acid sequences and predicted the corresponding protein structure of DbPsy using various bioinformatics tools. Finally, we presented some hints for the regulation mechanisms of DbPsy at the molecular level according to the computed results.

2 METHODS

2.1 Strains and culture conditions

Dunaliella bardawil cells, obtained from the Institute of Hydrobiology, Chinese Academy of Science, were grown in defined medium (Sheffer and Avron, 1986) containing 2 mol/l NaCl at 26°C under a 14/10 h dark/light cycle, and were collected at the log phase or late log phase. *Escherichia coli* TOP10 was used as the host for the multiplication of plasmids.

2.2 Known sequence verification

For validation of the *psy* sequence deposited at NCBI (GenBank: EU328288.1), two pairs of overlapping primers (Table 1) were designed according to the Psy mRNA sequence (GenBank: EU328287.1).

Total RNA was extracted from 6 ml of *D.bardawil* cells grown at the late log phase using E.Z.N.A. Total RNA Kit II (OMEGA) following conditions recommended by the manufacturer. First strand cDNA was synthesized with Oligo dT-Adaptor Primer in a total volume of 10 µl, and subsequent PCR amplification of two cDNA fragments, with two pairs of specific primes (Dbpsy-F and Dbpsy-inR; Dbpsy-inF and Dbpsy-R), respectively, was fulfilled using an RNA PCR Kit (AMV) ver. 3.0 (TaKaRa) according to the manufacturer's protocol. The PCR procedure to amplify the Psy cDNA fragments is as follows: 95°C, 5 min; 35 cycles of 94°C, 45 s; 60°C, 45 s; and 72°C, 1 min. The amplified fragments were cloned to pMD 19-T vector (TaKaRa) and sequenced.

2.3 Genomic walking for isolation of *psy* promoter and terminator

Based on the sequences of the *psy* gene of *D.bardawil* (GenBank: EU EU328288.1), three gene-specific primers for cloning of *psy* promoter, pSP1, pSP2 and pSP3, were designed (Table 1). Likewise, the primers for isolation of *psy* terminator, tSP1, tSP2 and tSP3, are also shown in Table 1. Genomic DNA of *D.bardawil* was extracted following the method described by Yang *et al.* (2000). Using Genomic Walking Kit (TaKaRa), LA PCR-based genomic walking was performed to obtain the *psy* promoter and terminator, respectively. The primary nested PCR products were diluted to 1:100 with distilled water for subsequent nested PCR reactions; all manipulations were corresponding with the manufacturer's protocol. Validation of putative *psy* promoter and terminator were performed using primers Ptest For (Table 1) and pSP3 or primers tSP3 and Ttest Rev (Table 1), respectively. Sequence analysis was performed using Blast Software (http://blast.ncbi.nlm.nih.gov/), promoter predictions software PlantPAN (http://plantpan.mbc.nctu.edu.tw/), CorePromoter program (http://rulai.cshl.org/tools/genefinder/CPROMOTER/) for transcriptional start site (TSS) search and terminator scan program Poly (A) Signal Miner (http://dnafsminer.bic.nus.edu.sg/) and RibEx (http://132.248.32.45/cgi-bin/ribex.cgi).

Table 1. Primers used in this study ($5' \rightarrow 3'$)

Procedure	Primer	Primer sequence
PCR	Dbpsy-F	ATGGCACAGCGAACAGCAACTT CCTCCTCCTCCTCTCCTAGCATCA
	Dbpsy-inR	GC***CTGAGGCGTGATCTTTGA***
	Dbpsy-inF	GCA***TCAAAGATCACGCCTCAG***
	Dbpsy-R	TTATTTGTTCTTGGGCACCAAG
Genomic walking	pSP1	CGAGGTAGAAGGTCTTGGCGTA
	pSP2	GGCACTGACTGAGCGGTCTTTA
	pSP3	CTGCTCTGCGACTACCATTCCT
	tSP1	AATGCCCAAAGCAGAAGACAGG
	tSP2	CACACTACCCTTGGTGATGTCTTG
	tSP3	AGACTGGCTGCCGTTTCCTGAT
Sequence validation	Ptest For	AACAATGCCACCAATACC
	Ttest Rev	CGACGGATGACTGATTACGC

Dbpsy-inR and Dbpsy-inF are overlapping primers. Complement sequences of the DbPsy cDNA are in bold and italic.

2.4 Bioinformatics analysis and phylogenetic construction

Psy splicing pattern was predicted through NCBI Splign (http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi?textpage=online&level=form) and GENSCAN (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=genscan). Component analysis of *psy* was calculated using DNAStar software 7.1.0. Physical and chemical characteristics of DbPsy were analyzed by ProtParam tool (http://expasy.org/tools/protparam.html). Subcellular localization presumption was performed using WoLF PSORT (http://wolfpsort.org/). Conserved domains in DbPsy were detected using the Conserved Domains Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Secondary structure was predicted via the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/); 3D structure was constructed using 3D-JIGSAW (http://bmm.cancerresearchuk.org/~3djigsaw/). Multiple alignments among similar enzymes were conducted using Clustal X 1.83. Phylogenetic and molecular evolutionary analysis of the amino acid sequences of different Psys were conducted using the Neighbor Joining method and the molecular evolution genetics analysis (MEGA) software, version 4.0.2.

3 RESULTS

3.1 Isolation and sequence analysis of the full-length DbPsy cDNA

The Psy template deposited at NCBI comes from direct submission without any support of report hitherto, which necessitates the validation work before genomic walking. To this end, RT–PCR

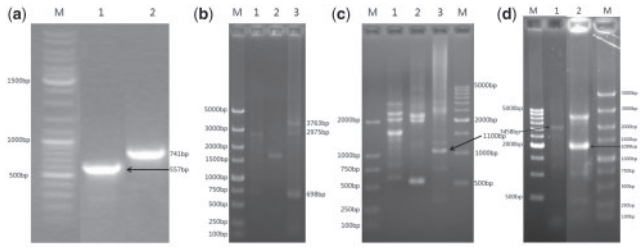

Fig. 1. Isolation of promoter and terminator of *psy* gene in *D.bardawil*. (**a**) Validation of *psy* template. M, 100 bp DNA marker; land 1 and 2, Psy cDNA fragments with 21 bp overlapped. (**b**) Promoter isolation by LA PCR-based genome walking. Three fragments were obtained and sequenced; the shortest one with 689 bp nucleotides did not have any homology with the upstream of *psy*, while the 3763 bp fragement was supposed probably to be the 5′-UTR of *psy*. M, DL5000 DNA marker; land 1–3, products of 1st nested PCR, 2nd nested PCR and 3rd nested PCR, correspondingly. (**c**) Isolated fragment assumed to be the terminator of *psy*. The brightest fragment of 1100 bp recovered, with an 184 bp region identical to the downstream of *psy*, was inferred to be the *psy* terminator. M, DL2000 DNA marker; lane 1–3, products of 1st nested PCR, 2nd nested PCR and 3rd nested PCR, respectively; M, 500 bp DNA marker. (**d**) Validation of hypothetic *psy* promoter and terminator using testing primers Ptest For and Test Rev with pSP3 and tSP3, respectively. Two segments, length in 3458 bp and 1099 bp, respectively, corresponding to *psy* promoter and terminator, were obtained.

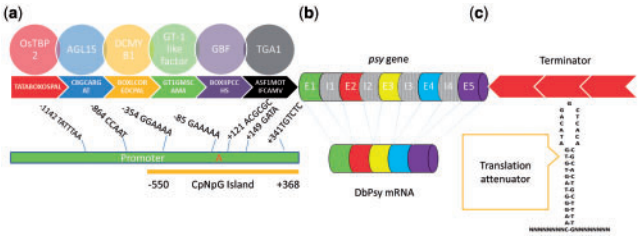

Fig. 2. Structure of full-length genomic DNA of *psy* including promoter and terminator. (**a**) A 3633 bp length promoter was obtained and analyzed. Several conserved motifs in addition to a CpNpG Island were identified in the promoter: (1) TATABOXOSPAL, a typical box found in various species (Zhu *et al*., 2002); (2) C8GARGAT, a A/T-rich binding site for plant MADS-domain protein AGL15 (Tang and Perry, 2003); (3) BOXLCOREDCPAL, box-L-like sequences with consensus ACCTTCC core sequences for a transcriptional activator DCMYB1 (Maeda *et al*., 2005); (4) GT-1GMSCAM4, GT-1 motif found in the promoter of soybean SCaM-4, whose expression is induced by pathogen and salt through GT-1 like factor (Park *et al*., 2004); (5) BOXIIPCCHS, Box II/G box found in the parsley *chs* genes, essential for light regulation (Block *et al*., 1990); (6) ASF1MOTIFCAMV, TGACG motif for the basic domain/leucine zipper transcription factor TGA1, may be relevant to light regulation (Després *et al*., 2003). (**b**) *Psy* gene consists of six exons separated by five introns franked by canonical GT donor site and AG acceptor site. (**c**) Though no poly (A) signal was detected, we predicted a reverse translation attenuator inside the last intron.

was performed and two cDNA fragments, 555 bp and 741 bp in length respectively, were obtained with overlapping primers Dbpsy-F/inR and Dbpsy-inF/R, correspondingly (Fig. 1a). Sequencing and sequence assembly results fed back a complete full-length cDNA of DbPsy (GenBank: EU328287.1, 1275 bp in length), which validates the *psy* gene deposited at NCBI (GenBank: EU328288.1) indirectly.

3.2 Isolation of the promoter and terminator of DbPsy

Three candidate promoter fragments of the third nested PCR were purified and sequenced (Fig. 1b). Sequence analysis by BlastN found that both the 3763 bp and 2975 bp fragments possess identical regions (130 bp in length) as expected with the 5′-end of *psy*. Comparison of the sequences showed that the 3763 bp fragment is the extension of the shorter one. Likewise, a putative terminator fragment shares 184 bp identical nucleotides with the 3′-end of *psy* as intended (Fig. 1c). The AGE and sequencing results further validate the hypothetic promoter and terminator (Fig. 1d). These results suggested that the obtained sequences are the upstream and downstream UTR of *psy*, correspondingly.

3.3 Computational analysis of promoter and terminator of DbPsy

BlastN search found that the obtained sequences have great homology with the 5′- and 3′-end of *psy* gene, respectively. CorePromoter detection suggested that the TSS, 'A' marked in Figure 2a, locates at 370 nt upstream of the initiator codon ATG.

Five species (Arabidopsis, Maize, Rice, Tomato and other) were selected when conducting transcriptional factor binding sites search at PlantPAN server (Chang *et al.*, 2008). The retrieval results showed that the isolated promoter possesses many conserved motifs of plant (Fig. 2a), e.g. GATA box at position +149, CCAAT boxes widely found in eukaryotic genes, CGCG boxes with conserved sequences GCGCGC and ACGCGC at position −344 and +121 correspondingly. In addition, some regulatory elements were also found, such as AFRAT involving in IAA regulation with consensus sequence TGTCTC at position +341 (Goda *et al.*, 2004), the light-regulated elements G box (position −28) and ASF-1 binding site (position +102), etc. Interestingly, two remarkable motifs were also found in the promoter: box-L-like core sequences (position −491, also named BOXLCOREDCPAL in Fig. 2a) and GT-1 motifs (position −85 and −2606). Previous researches has implied a valuable application of UV irradiation for accumulating carotenoids (Mogedas *et al.*, 2009; Salguero *et al.*, 2005; White and Jahnke, 2002; Xue *et al.*, 2005). Maeda *et al.* (2005) showed that BOXLCOREDCPAL involved in UV-B irradiation regulation in *Daucus carota*, therefore, such motif in *D.bardawil* might function in a similar way. The GT-1 box (also called GT1GMSCAM4, Fig. 2a) might contribute to pathogen- and NaCl-induced expression of the promoter (Park *et al.*, 2004).

Both the Signal Miner and RibEx programs (Abreu-Goodger and Merino, 2005) failed to detect any canonical poly (A) signal in the hypothetic terminator. While a reverse translation attenuator inside the last intron of *psy* was identified by RibEx at 322 bp upstream of the TAA terminator codon, as is shown in Figure 2c. Such disappointing results prompted us to clone and analyze *psy* terminator sequence again, resulting in the same outcome (data not shown). These results suggest that we may fail to isolate the genuine terminator of *psy* in *D.bardawil*, or the *psy* terminator may possess no canonical poly (A) signal.

3.4 Analysis of the genomic structure of *psy*

To further validate and elucidate the genomic structure of *psy*, NCBI Splign (Kapustin *et al.*, 2008) and GENSCAN (Burge and Karlin, 1997) were performed. Both servers returned a consistent result that

Table 2. Components of *psy* gene in *D.bardawil*

	Start	End	Length (bp)	GC content (%)
Exon				
1	1	555	555	60.54
2	810	1070	261	55.17
3	2090	2245	156	56.41
4	2614	2765	152	55.92
5	3150	3300	151	49.67
Total			1275	52.85
Intron				
1	556	809	254	49.61
2	1071	2089	1019	47.2
3	2246	2613	368	53.8
4	2766	3149	384	54.95
Total			2025	50.33
mRNA	1	1275	1275	47.15

psy contains six exons separated by five introns (Fig. 2b). As shown in Table 2, the GC content (%) of all six exons varies from 49.67% to 60.54%, whereas the GC content (%) of the five introns ranges from 47.2% to 54.95%. Although the introns of *psy* are flanked by conventional 5′ splice donor GT and 3′ splice acceptor AG, they appear to be lack of A + T-rich region, which characterizes a canonical intron.

3.5 Physical and chemical characteristics of DbPsy

The 424 amino acid peptide possesses a computed molecular weight of 48.30 kDa, a theoretical isoelectric point (pI) of 9.16, and exhibits an aliphatic index of 76.65. It contains 46 negatively charged residues (Asp + Glu) and 57 positively charged residues (Arg + Lys) with several hydrophilic and hydrophobic regions, as determined by ExPASy (Gasteiger *et al.*, 2003). In addition, it is classified as an unstable protein with a considered methionine N-terminus. WoLF PSORT showed that DbPsy may be situated in the chloroplast (Horton *et al.*, 2007).

3.6 Conserved domain and motifs in DbPsy

Through the NCBI Conserved Domain Search (Marchler-Bauer *et al.*, 2009), the deduced amino acid sequence was specifically hit by the *trans*-Isoprenyl Diphosphate Synthases (Trans_IPPS) which included squalene and phytoene synthases. The prediction showed that DbPsy can be classified into the superfamily of IPPS and Class I terpene cyclases. Six regions together forming Trans_IPPS were also predicted in DbPsy (Fig. 3): substrate binding pocket, Mg^{2+} binding site, active site lid residues, catalytic residues and two aspartate-rich sites (171-DELVD-175, 297-DVGED-301). In rat, the catalytic site consists of a large central cavity formed by mostly antiparallel alpha helices with two aspartate-rich regions (DXXXD) (Gu *et al.*, 1998). The catalytic residues involved in the reactions of forming squalene. However, unlike squalene synthase, Psy does not require NADPH for its activity (Gu *et al.*, 1998). The two aspartate-rich motifs, along with Tyr and Phe rich in the rest conserved regions, may involve in condensation reactions in DbPsy (Tansey and Shechter, 2000).

3.7 The advanced structures of DbPsy

The secondary structure prediction run at the PSIpred server (McGuffin *et al.*, 2000) showed that the DbPsy consists of 72.64%

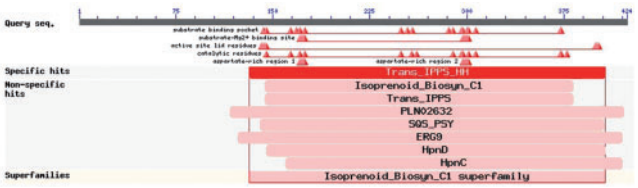

Fig. 3. Conserved domains in DbPsy detected by NCBI Conserved Domains Search. There is a conserved domain homologous with the *trans*-Isoprenyl Diphosphate Synthases in DbPsy from 132 to 407 amino acids, which consists of six motifs: substrate binding pocket, Mg^{2+} binding site, active site lid residues, catalytic residues and two aspartate-rich sites.

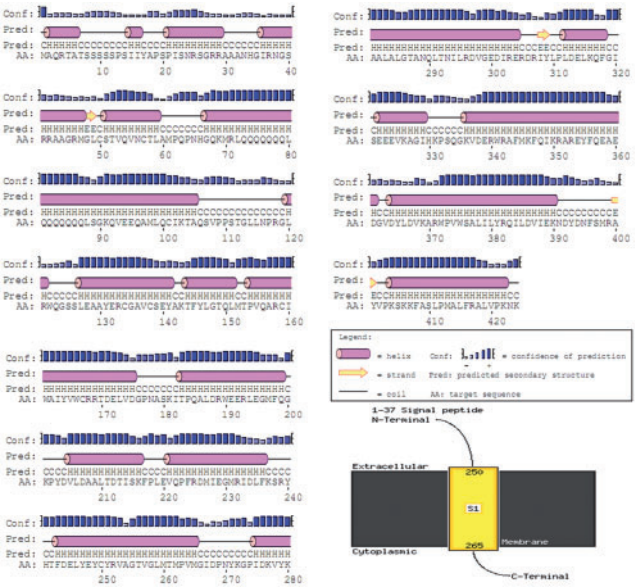

Fig. 4. Predicted secondary structure of DbPsy using the PSIPRED server. A transmembrane motif is revealed, along with a 37 amino acid signal peptide at the N-terminus extracellular (lower right).

α-helix (20 helices; 308 residues), 1.42% extended-beta (3 strands; 6 residues) and 25.94% random coil (23 coils; 110 residues) configurations (Fig. 4). MEMSAT3 and MEMSAT-SVM program detected a transmembrane region between amino acids 250 and 265. The carboxy-terminus is possible in the cytoplasm; while the amino-terminus, followed by a 37 amino acid signal peptide, locates in the extracellular fluid (Fig. 4).

Moreover, the tertiary structure of DbPsy was also constructed by 3D-JIGSAW (Contreras-Moreira and Bates, 2002). The calculated result shown in Figure 5 revealed that the single polypeptide chain model covers 322 residues from 103 to 424 amino acids, which is composed of 209 H-bonds, 20 helix elements and 32 turns. However, unlike the secondary structure retrieved from PSIPRE server, there is no strand element predicted using the 3D-JIGSAW server.

3.8 Homologous alignment and phylogenetic analysis of DbPsy

The gene family alignment of amino acidic sequences obtained by ClustalX analysis revealed that the overall structure of the

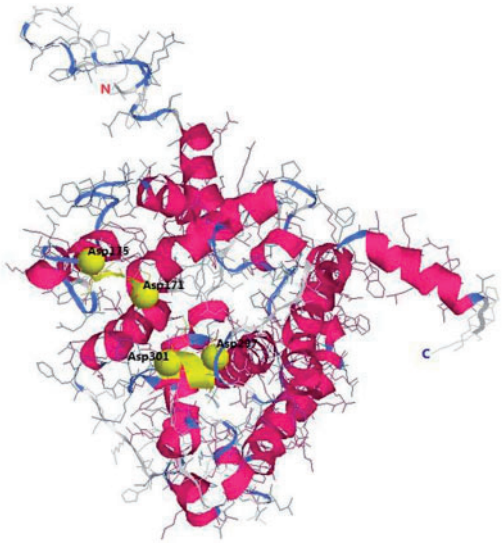

Fig. 5. The tertiary structure of DbPsy. Comparative modeling was performed using 3D-JIGSAW. The structure was presented as ribbons (the secondary structure elements) and wireframe using RasMol version 2.7.2.1 .1. The α-helix and β-turn regions of the putative protein were indicated with cylinders (red) and bands (blue), respectively. The two aspartate-rich motifs were colored by yellow, and the distinctive Asp residues (171, 175, 297 and 301) were labeled and highlighted as spacefill. The N-terminus and C-terminus were also shown.

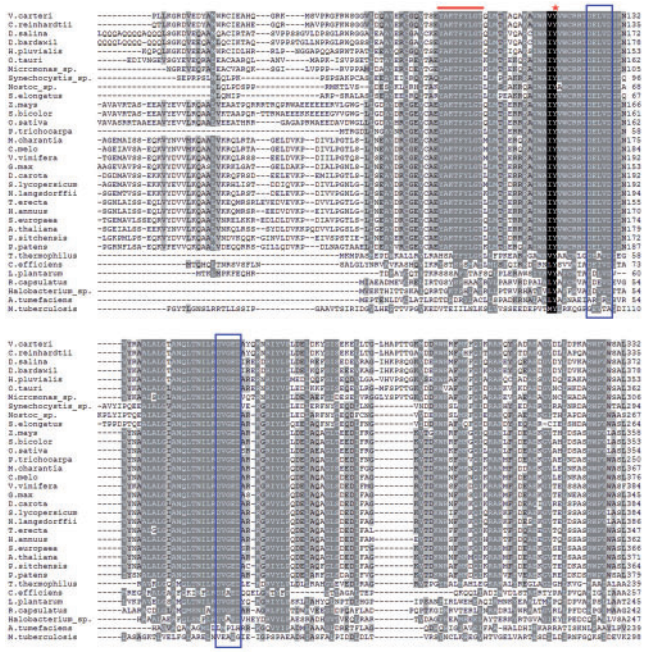

Fig. 6. The hypothetic Trans_IPPS_HH domain found in DbPsy. Comparison of deduced amino acid sequence of DbPsy with its homologs indicated a functional conserved domain Trans_IPPS_HH, which lays in the C-terminus from 132 to 424 amino acids. The active site lid residues (141 aa to 145 aa) are marked with red line above; the highly conserved catalytic residue Try164 is marked with red asterisk; two aspartate-rich regions are boxed in blue. Gaps introduced to maximize sequence homology are indicated by dashes. Amino acid residues identical in all sequences are highlighted in white type on a black background; less conserved residues appear on a light gray background. Psys are shown as GenBank accession number: AAT46069.1, AAW28851.1, XP_002956783.1, XP_001701192.1, CAL54118.1, XP_002508518.1, XP_001775752.1, ABB29857.1, XP_002302132.1, XP_002442578.1, ABR16198.1, AAR86104.1, ADC34069.1, ACU17983.1, AAX19898.1, XP_002271575.1, AAM45379.1, NP_001108117.1, ABU40771.1, ABA99494.1, NP_197225.1, ABB52068.1, CAC19567, NP_441168, NP_485873, NP_217974, YP_006040, P17056, NP_786525, AAL42564, NP_280449, NP_737251, CAA45350.

Trans_IPPS_HH domains has been conserved from bacteria to higher plant (Figs 3 and 6), though all of the homologs aligned varied considerably in their N-termini (data not shown). The functional conserved domain may cover from 132 to 424 amino acids.

The BlastP search demonstrated that DbPsy shares the highest homology with *D.salina* at the protein level (79% identity and 87% similarity), 68% identity and 80% similarity with the *Haematococcus pluvialis* counterpart, 69% and 81% with the *Chlammydomonas reinhardtii* counterpart, 55% and 69% with the *Arabidopsis thaliana* counterpart. A lower homology of DbPsy was found with the prokaryote counterparts, with the exception of cyanobacterial Psys (40–42% identity).

The phylogenetic tree was constructed using Neighbor Joining method by MEGA 4.0.2 software (Tamura *et al.*, 2007) (Fig. 7). Three defined clusters can be found: Clusters I, II and III. Cluster I contains Psys from higher plants, cluster II consists of the algae Psys and cluster III comprises the bacterial homologs. Such phylogenetic dendrogram showed similar patterns with BlastP search and DNA sequence alignment results (data not shown); which indicated that Psys isolated from green algae can be classified into one group (Fig. 7). Accordingly, these results indicated an evolutionary relationship among green algae.

4 DISCUSSION

Duanliella bardawil and *D.salina* are recognized as the most promising natural sources to produce β-carotene in biotechnology commercial-production applications, due to their ability of massive accumulation of β-carotene under stress conditions, such as high light intensity and salinity, stress temperature, privative nutrients, etc. Plenty of studies have been performed to investigate the carotenoid biosynthesis pathway for the ultimate goal to produce as much β-carotene as possible (Mogedas *et al.*, 2009; Salguero *et al.*, 2005; Xue *et al.*, 2005). Nevertheless, most of these researches mainly focused on culture conditions and extraction approaches, or other aspects except carotenogenic enzyme genes. So far, a small number of enzymes directly involving in carotenoid biosynthetic pathway have been studied in *Dunaliella*, especially in *D.bardawil* (Yan *et al.*, 2005; Zhu *et al.*, 2005). Psy is one of the few which were cloned in *D.salina* and *D.bardawil*. However, it has not yet been experienced deeply functional validation up to now in *D.bardawil*. The attempts to obtain *psy* regulatory elements from *D.bardawil* would facilitate interpretation of the regulation mechanisms for Psy expression and accumulation of β-carotene. Additionally, works on elaborating the Psy structure at the protein level would help further presentation of the differentiations of Psy regulation even the

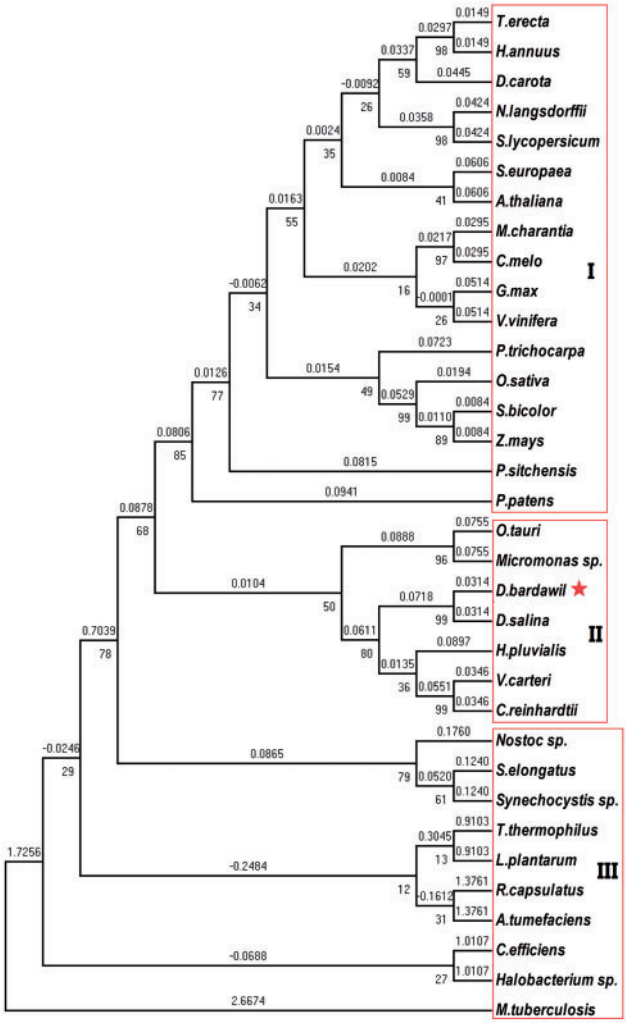

Fig. 7. Phylogenetic relationship of the DbPsy to other Psys of various species. The Psys used for phylogenetic tree construction are as the amino acid sequences used for homology alignment above. Numbers below the branches are the Neighbor Joining bootstrap values. The evolutionary distance is reflected by the numbers above the branches and the branch lengths proportional to the degree of amino acid substitutions.

whole carotenoid metabolic pathway between species. We strongly speculate, based on previous intensive efforts (Ben-Amotz *et al.*, 1987; León *et al.*, 2005; White and Jahnke, 2002; Xue *et al.*, 2005), that regulation of Psy and carotenogenesis is fulfilled at multiple levels from DNA structure to post-translational control. More specifically, the regulatory elements of *psy* gene and its protein structure may confer different regulation mechanisms in *D.bardawil* from other aglae.

As shown previously, massive accumulation of β-carotene is triggered by environmental stress (Mogedas *et al.*, 2009; White and Jahnke, 2002; Ye *et al.*, 2008), implying a regulatory expression pattern of many carotenogenic genes. Among these environmental stresses, UV interference executes different effects on β-carotene accumulation through photosynthetic responses (Mogedas *et al.*, 2009; White and Jahnke, 2002). One of the protective mechanisms against UV-induced photosynthetic damage involves in massive accumulation of β-carotene in *D.bardawil* exposed to UV-A rather than UV-B, while *D.salina* lacks such ability (White and Jahnke, 2002). UV-B even executes contrasting effects on β-carotene biosynthesis: decreases β-carotene but enriches other carotenoids including phytoene. However, this UV-A-induced protection is lost in cells treated with the bleaching herbicide norflurazon (Ben-Amotz *et al.*, 1987). In such cells phytoene synthesis is enhanced significantly, with a decreased level of β-carotene, so that León *et al.* (2005) applied this bleaching herbicide to product phytoene in *D.bardawil*. Our study found that the promoter of DbPsy possesses a conserved motif BOXLCOREDCPAL (Fig. 3) homologous with DcPAL1 counterpart in *D. carota*. Such motif participates in up-regulation responses of DcPAL1 to UV-B irradiation in *D. carota* mediated by DcMYB1 (Maeda *et al.*, 2005). These proofs suggest that, in accordance with the investigations discussed above (Ben-Amotz *et al.*, 1987; Mogedas *et al.*, 2009; White and Jahnke, 2002), DbPsy may be regulated by UV-B and norflurazon with respect to similar photobleaching effects to accumulate phytoene, rather than β-carotene, through its promoter region. We conclude that UV-B and norglurzon may induce photobleaching through the BOXLCOREDCPAL situated within the promoter of DbPsy.

Moreover, a salt- and light-induced pattern of β-carotene should be concerned with certain regulatory domains in carotenogenic genes. Here, we found two GT-1-like boxes deposited inside the isolated promoter, which were widely discovered in many light-regulated genes (Terzaghi and Cashmore, 1995). With respect to GT1CONSENSUS (GAAAAA, position −354), we speculate an intermediary role on interaction between TFIIA and GT-1-like factors in *D.bardawil* (Terzaghi and Cashmore, 1995): TFIIA indirectly interacts with GT-1-like factors through GT1CONSENSUS to accumulate β-carotene induced by light. Another GT-1-like box (Fig. 3, GT1GMSCAM4) involves in NaCl-induced up-regulated expression of SCaM-4 in Glycine max (Park *et al.*, 2004), consequently, GT1GMSCAM4 found in DbPsy promoter may fulfill similar function as its counterpart in Glycine max: up-regulates its host gene *psy* in order to enhance β-carotene synthesis by NaCl. Since both motifs up-regulate their corresponding genes in their own species, we suppose parallel effect in *D.bardawil*. Furthermore, as described above, the contrasting effects of UV-A and UV-B exclusively in *D.bardawil* (White and Jahnke, 2002) suggest UV-B may also up-regulate the expression of Psy through these motifs. Such unique characteristics imply fine tuning of Psy (even the whole carotenogenesis) related to diversity of DNA-protein interactions between *D.bardawil* and other algae.

Psy has been regarded as the first key enzyme to catalyze a head to head combination reaction, generating colorless phytoene in carotenoid biosynthesis. It is supposed to be the regulatory point controlling the flux of carbon source towards carotenoids (Shewmarker *et al.*, 1999). To further illuminate the regulation mechanisms of DbPsy, we performed a series of *in silico* analysis at the protein level. Advanced protein structure analysis manifests a relatively conserved C-terminus containing putative substrate binding and catalytic domains from around 100 to 400 amino acids. Phylogenetic analysis indicated that Psys isolated from algae occupy one cluster (Fig. 7). These clues imply that they may share some similar regulation mechanisms distinctively with regard to the C-terminus of Psy in algae. Furthermore, alignments of the deduced amino acid sequence showed the sequence diversity

between species primarily in the N-terminus. Tran *et al.* (2009) suggested the N-terminal regions, unessential to the enzymatic function, uniquely account for the major differences of the two Psy classes found in some algae. Moreover, the N-terminal region of plant Psy is generally longer than bacterial enzymes (Kim *et al.*, 2003). This region is crucial for protein targeting and dimerization that are critical for enzyme stability and activity (Cunningham and Gantt, 1998; Sun *et al.*, 1996). The N-terminus diversity between *D.bardawil* and other algae may affect Psy turnover and catalytic activity in certain similar patterns. In addition, increased level of Psy does not directly correlate with increased carotenoids content (Kim *et al.*, 2003); the N-terminal variety may render different capabilities between different Psy members of inter and intra species with regard to protein targeting, processing or assembly into a fully functional complex. As discussed above, we speculate the sequence differences of Psy at the protein and DNA levels may be responsible for the fine tuning differences between *D.bardawil* and *D.salina,* we also conducted alignments between these two algae, which returned similar diversity of N-terminus (or the 5′-end of the Psy mRNA CDS) and conserved C-terminus (or the 3′-end of the Psy mRNA CDS) (data not shown). Accordingly, we speculate that the fine tuning mechanisms of Psys between algae may lie in the variation of their N-terminus. Studies on the N-terminal diversity as well as other differential regions between species would be helpful to clarify the distinct regulation mechanisms for Psy in *D.bardawil*.

An in-depth understanding of the regulation mechanisms of DbPsy still requires further intensive investigations on the promoter function involving in the putative motifs discussed above and other related conserved sequences. Moreover, the identification of *psy* promoter in *D.salina* and subsequent comparison with the one in *D.bardawil* will shed light on the fine tuning differences between the two algae at the DNA level. Besides, the protein structure is also a key point to aid the presentation of these mechanisms, especially its N-terminus. Such understanding would also contribute to bioengineering using DbPsy promoter to product foreign proteins.

Funding: National Natural Foundation of China (grant 30870025).

Conflict of Interest: none declared.

REFERENCES

Abreu-Goodger,C. and Merino,E. (2005) RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acid Res.*, **33**, W690–W692.

Ben-Amotz,A. *et al*. (1987) Massive accumulation of phytoene induced by norflurazon in *Dunaliella bardawil* (Chlorophyceae) prevents recovery from photoinhibition. *J. Phycol.*, **23**, 176–181.

Ben-Amotz,A. *et al*. (1982) Accumulation of b-carotene in halotolerant algae: purification and characterization of b-carotene-rich globules from *Dunaliella bardawil* (Chlorophyceae). *J. Phycol.*, **18**, 529–537.

Block,A. *et al*. (1990) Functional borders, genetic fine structure, and distance requirements of cis elements mediating light responsiveness of the parsley chalcone synthase promoter. *Proc. Natl Acad. Sci. USA*, **87**, 5387–5391.

Burge,C. and Karlin,S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, **268**, 78–94.

Chang,W.C. *et al*. (2008) PlantPAN: Plant Promoter Analysis Navigator, for identifying combinatorial *cis*-regulatory elements with distance constraint in plant gene group. *BMC Genomics*, **9**, 561–574.

Contreras-Moreira,B. and Bates,P.A. (2002) Domain Fishing: a first step in protein comparative modelling. *Bioinformatics*, **18**, 1141–1142.

Cunningham,F.X. Jr. and Gantt,E. (1998) Genes and enzymes of carotenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 557–583.

Demmig-Adams,B. and Adams,W.W. (1996) The role of xanthophyll cycle carotenoids in the protection of photosynthesis. *Trends Plant Sci.*, **1**, 21–26.

Després,C. *et al*. (2003) The *Arabidopsis* NPR1 disease resistance protein is a novel cofactor that confers redox regulation of DNA binding activity to the basic domain/leucine zipper transcription factor TGA1. *Plant Cell*, **15**, 2181–2191.

Farré,G. *et al*. (2010) Travel advice on the road to carotenoids in plants. *Plant Sci.*, **179**, 28–48.

Feltl,L. *et al*. (2005) Reliability of carotenoid analyses: a review. *Curr. Anal. Chem.*, **1**, 93–102.

Gasteiger,E. *et al*. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acid Res.*, **31**, 3784–3788.

Goda,H. *et al*. (2004) Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in Arabidopsis. *Plant Physiol.*, **134**, 1–19.

Goodwin,T.W. (1980) *The Biochemistry of Carotenoids*, vol. I, Plants. Chapman & Hall, London, p. 98.

Gu,P. *et al*. (1998) Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase. *J. Biol. Chem.*, **273**, 12515–12525.

Horton,P. *et al*. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acid Res.*, **35**, W585–W587.

Kapustin,Y. *et al*. (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol. Direct*, **3**, 20–32.

Kim,I.J. *et al*. (2003) Expression and activity of citrus phytoene synthase and β-carotene hydroxylase in *Escherichia coli*. *J. Microbiol.*, **41**, 212–218.

Landrum,J.T. and Bone,A.R. (2001) Lutein, zeaxanthin, and the macular pigment. *Arch. Biochem. Biophys.*, **385**, 28–40.

León,R. *et al*. (2005) Production of phytoene by herbicide-treated microalgae *Dunaliella bardawil* in two-phase systems. *Biotechnol. Bioeng.*, **92**, 695–701.

Maeda,K. *et al*. (2005) DcMYB1 acts as a transcriptional activator of the carrot phenylalanine ammonia-lyase gene (DcPAL1) in response to elicitor treatment, UV-B irradiation and the dilution effect. *Plant Mol. Biol.*, **59**, 739–752.

Marchler-Bauer,A. *et al*. (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acid Res.*, **37**, D205–D210.

McGuffin,L.J. *et al*. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.

Michaud,D.S. *et al*. (2000) Intake of specific carotenoids and risk of lung cancer in 2 prospective US cohorts. *Am. J. Clin. Nut.*, **72**, 990–997.

Mogedas,B. *et al*. (2009) Beta-carotene production enhancement by UV-A radiation in *Dunaliella bardawil* cultivated in laboratory reactors. *J. Biosci. Bioeng.*, **108**, 47–51.

Muthukannan,P. *et al*. (2010) *In vitro* evaluation of β-carotene production in two different strains of *Dunaliella salina* Teodoresco (Chlorophyta). *J. Biosci. Res.*, **1**, 83–87.

Olson,J.A. 1994. Needs and sources of carotenoids and vitamin. *Nutr. Rev.*, **52**, S67–S73.

Park,H.C. *et al*. (2004) Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. *Plant Physiol.*, **135**, 2150–2161.

Salguero,A. *et al*. (2005) UV-A mediated induction of carotenoid accumulation in *Dunaliella bardawil* with retention of cell viability. *Appl. Microbiol. Biotechnol.*, **66**, 506–511.

Salvini,M. *et al*. (2005) cDNA cloning and expression of the phytoene synthase gene in sunflower. *J. Plant Phys.*, **162**, 479–484.

Shaish,A. *et al*. (2006) 9-cis β-carotene-rich powder of the alga *Dunaliella bardawil* increases plasma HDL-cholesterol in fibrate-treated patients. *Atherosclerosis*, **189**, 215–221.

Sheffer,M. and Avron,M. (1986) Isolation of the plasma-membrane of the halotolerant alga *Dunaliella salina* using sulfur hodamine B as a probe. *Biochim. Biophys. Acta.*, **857**, 155–164.

Shewmarker,C.K. *et al*. (1999) Seed specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *Plant J.*, **20**, 401–412.

Siefermann-Harms,D. (1987) The light harvesting and protective functions of carotenoids in photosynthetic membranes. *Physiol. Plantarum*, **69**, 561–568.

Sun,Z. *et al*. (1996) Cloning and functional analysis of the β-carotene hydroxylase of *Arabidopsis thaliana*. *J. Biol. Chem*, **271**, 24349–24352.

Tamura,K. *et al*. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, **24**, 1596–1599.

Tang,W. and Perry,S.E. (2003) Binding site selection for the plant MADS-domain protein AGL15: an in vitro and in vivo study. *J. Biol. Chem.*, **278**, 28154–28159.

Tansey,T.R. and Shechter,I. (2000) Structure and regulation of mammalian squalene synthase. *Biochim. Biophys. Acta*, **1529**, 49–62.

Terzaghi,W.B. and Cashmore,A.R. (1995) Light-regulated transcription. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **46**, 445–474.

Tran,D. *et al*. (2009) An update on carotenoid biosynthesis in algae: phylogenetic evidence for the existence of two classes of phytoene synthase. *Planta*, **229**, 723–729.

White,A.L. and Jahnke,L.S. (2002) Contrasting effects of UV-A and UV-B on photosynthesis and photoprotection of beta-carotene in two *Dunaliella* spp. *Plant Cell Physiol.*, **43**, 877–884.

Xue,L. *et al*. (2005) Effects of enhanced ultraviolet-B radiation on algae and cyanobacteria. *Crit. Rev. Microbiol.*, **31**, 79–89.

Yan,Y. *et al*. (2005) Cloning and sequence analysis of the phytoene synthase gene from a unicellular chlorophyte, *Dunaliella sanlia*. *J. Agric. Food Chem.*, **53**, 1466–1469.

Yang,Z.Y. *et al*. (2000) Construction of a genomic DNA library of *Dunaliella salina*. *Acta Phytophysiol. Sinica*, **26**, 25–28.

Ye,Z.W. *et al*. (2008) Biosynthesis and regulation of carotenoids in *Dunaliella*: progresses and prospects. *Biotech Adv.*, **26**, 352–360.

Zhu,Q. *et al*. (2002) Rice TATA binding protein interacts functionally with transcription factor IIB and the RF2a bZIP transcriptional activator in an enhanced plant in vitro transcription system. *Plant Cell*, **14**, 795–803.

Zhu,Y.H. *et al*. (2005) Isolation and characterization of phytoene desaturase cDNA involved in the β-carotene biosynthetic pathway in *Dunaliella salina*. *J. Agric. Food Chem.*, **53**, 5593–5597.