

文章编号: 1674-5566(2017)02-0171-12

DOI:10.12024/jsou.20160401744

普通小球藻八氢番茄红素合成酶基因 *psy* 的 cDNA 克隆及生物信息学分析

刘冬雨¹, 陶贤继¹, 魏 华²

(1. 上海海洋大学 水产与生命学院, 上海 201306; 2. 上海农林职业技术学院, 上海 201699)

摘要: 通过 RACE-PCR 首次克隆获得了普通小球藻 (*Chlorella vulgaris*) 中八氢番茄红素合成酶编码基因 *psy* 的 cDNA 全长序列。该序列全长 1 605 bp, 包括 1 245 bp 开放阅读框, 编码 415 个氨基酸。氨基酸序列对比及系统发生分析表明其与其他物种的 PSY 蛋白具有同源性, 与绿藻聚为一支, 与小球藻 (*Chlorella variabilis*)、原壳小球藻 (*Auxenochlorella protothecoides*) 和佐夫色绿藻 (*Chromochloris zofingiensis*) 的遗传距离最近 (分别为 0.173、0.188 和 0.239), 并具有较高的相似性 (81% ~ 84%) 和一致性 (69% ~ 76%)。亚细胞定位预测表明普通小球藻 PSY 前 45 个氨基酸残基为叶绿体转运肽, 可能引导翻译后的肽链进入叶绿体。保守区块、特征基序分析及三维建模显示, 序列中具有多个潜在的蛋白质修饰位点, 以及 PSY 蛋白的保守区块和特征基序, 包括两个保守的底物-镁离子结合位点和两个活性位点遮蔽残基, 用于结合底物及保护催化反应中间产物; 普通小球藻 PSY 蛋白中还具有一个特异的底物-镁离子结合位点, 其活性和功能还未知。总之, 研究结果揭示了普通小球藻 *psy* 基因的 cDNA 全长序列及可能的蛋白质序列结构特性, 结果可为构建过表达载体, 提高类胡萝卜素产量提供相关信息。

关键词: 八氢番茄红素合成酶; *psy* 基因; 普通小球藻; 类胡萝卜素

中图分类号: S 917 **文献标志码:** A

类胡萝卜素是一类存在于高等植物、微藻和部分光合细菌中的天然色素, 由于其抗氧化特性^[1], 使其对包括癌症在内的一些疾病具有预防作用^[2-5]。对类胡萝卜素天然合成过程的研究有助于提高产量, 满足不断提升的市场需求。在类胡萝卜素生物合成途径中, 八氢番茄红素合成酶 (phytoene synthase, PSY) 是催化关键步骤的限速酶, 能影响整个类胡萝卜素合成的反应速率^[6-7]。因此, 深入研究 PSY 的编码基因序列 (*psy*)、酶结构及功能, 有助于改进类胡萝卜素合成调控, 或通过基因工程手段提高类胡萝卜素产物合成。

微藻中含有许多特有的类胡萝卜素^[8], 同时易于培养, 因而成为类胡萝卜素自然来源的首选。近年来, 雨生红球藻 (*Haematococcus pluvialis*)^[9]、佐夫色绿藻 (*Chromochloris*

zofingiensis)^[10]、原壳小球藻 (*Auxenochlorella protothecoides*)^[11] 等藻类的 *psy* 基因全长序列相继被克隆鉴定。通过构建过表达载体及细胞核转化, 成功将佐夫色绿藻的 *psy* 基因转入莱茵衣藻 (*Chlamydomonas reinhardtii*) 中, 使细胞内紫黄质 (Violaxanthin) 和叶黄素 (Lutein) 的含量上升^[10]。

小球藻 (*Chlorella*) 富含重要的类胡萝卜素^[12], 其中叶黄素能够增强鲤体内抗氧化酶的活性, 从而增强抵抗有害物质的能力^[13]。普通小球藻 (*Chlorella vulgaris*) 作为小球藻属的典型种, 易于培养、生长迅速, 结构简单便于研究, 在藻类研究中常作为模式生物, 也是天然类胡萝卜素的重要提取源^[14]。因此, 本文选取普通小球藻作为研究对象, 首先克隆测定其 *psy* 基因的 cDNA 全长

收稿日期: 2016-04-18 修回日期: 2016-09-03

基金项目: 国家自然科学基金 (41272381); 上海市知识服务平台上海海洋大学水产动物遗传与育种中心项目 (ZF1206); 国家海洋局公益项目 (201505034)

作者简介: 刘冬雨 (1991—), 男, 硕士研究生, 研究方向为生物学。E-mail: dyl.losteva@aliyun.com

通信作者: 魏 华, E-mail: weih@shafc.edu.cn

序列,并对其编码蛋白质的序列进行预测,分析其同源性并构建系统发生树。同时预测 PSY 蛋白质的基本理化性质、功能位点和亚细胞定位,并构建蛋白质高级结构的三维模型。

1 材料与方法

1.1 藻种培养方法

普通小球藻(*C. vulgaris*)购于中国科学院淡水藻种库(武汉)。藻细胞培养于1 000 mL三角烧瓶中,使用BG-11培养液^[15],pH为7~8,藻液总体积400 mL。使用光照培养箱维持稳定的培养条件,温度(25±1)℃,光照强度8 000 lx,光照周期为12 h(09:00至21:00):12 h。为获得足够的总RNA,当藻细胞生长进入指数期末期(第7天,21:00)时收集藻细胞,用于提取总RNA及*psy*克隆。

1.2 总RNA提取

使用TRIzol试剂(Invitrogen)按照使用指南提取总RNA。总RNA的浓度及纯度使用Nanodrop2000(Thermo Scientific)测定,根据OD₂₆₀计算RNA浓度(RNA浓度=OD₂₆₀×40 ng/μL),根据OD₂₆₀/OD₂₈₀及OD₂₆₀/OD₂₃₀判定RNA的纯度。用RNase-Free水将总RNA稀释至终浓度为500 ng/μL,保存于-80℃备用。总RNA的完整性使用1%琼脂糖凝胶电泳测定。

1.3 *psy* cDNA全长的克隆

以总RNA为模板,使用反转录酶M-MLV(TaKaRa)及引物Oligo(dT18,TaKaRa)合成普通cDNA第一链。使用SMARTer PCR cDNA Synthesis Kit(Clontech)合成3'RACE及5'RACE cDNA。操作过程均按试剂说明书。

*psy*保守片段的克隆:根据GenBank(<https://www.ncbi.nlm.nih.gov/genbank/>)数据库中已登录的高等植物和微藻的*psy*基因核酸序列及氨基酸序列,使用Clustal Omega(<http://www.ebi.ac.uk/Tools/msa/clustalo/>)进行多序列比对。核酸序列通过GeneFisher2(<http://bibiserv.techfak.uni-bielefeld.de/genefisher2/>)生成简并序列,并用Oligo7(7.6,Molecular Biology Insights)生成简并引物;氨基酸序列通过Codehop(<http://blocks.fhcr.org/codehop.html>)处理生成简并引物。综合以上两种方法,设计一对简并引物(*psy*-pf和*psy*-pr,表1)用于第一轮PCR扩增,第二对非简

并引物(*psy*-yf和*psy*-yr,表1)用于第二轮PCR扩增。PCR使用GoTaq Green Master Mix试剂(Promega)。第一轮以普通cDNA为模板,扩增程序:95℃预变性5 min,40个PCR循环(95℃变性30 s,47℃退火30 s,72℃延伸1 min),及72℃延伸5 min。PCR产物稀释50倍后作为第二轮PCR模板,扩增程序:95℃预变性5 min,20个降落式PCR循环(95℃变性30 s;退火温度由70℃降落至61℃,每两个循环降低1℃,退火30 s;72℃延伸1 min),25个PCR循环(95℃变性30 s,60℃退火30 s,72℃延伸1 min),及72℃延伸5 min。

3'末端片段的克隆:第一轮PCR以3'RACE cDNA为模板,上游引物*psy*-3GSP-Fo,下游引物NUP,扩增程序:95℃预变性5 min,40个PCR循环(95℃变性30 s,64℃退火40 s,72℃延伸2 min),及72℃延伸5 min。PCR产物稀释50倍后作为第二轮PCR模板,上游引物替换为*psy*-3GSP-Fi,扩增程序:95℃预变性5 min,5个PCR循环1(95℃变性30 s,72℃退火延伸2 min),5个PCR循环2(95℃变性30 s,70℃退火40 s,72℃延伸2 min),5个PCR循环3(95℃变性30 s,68℃退火40 s,72℃延伸2 min),5个PCR循环4(95℃变性30 s,67℃退火40 s,72℃延伸2 min),30个PCR循环5(95℃变性30 s,66℃退火40 s,72℃延伸2 min),及72℃延伸5 min。

5'末端片段的克隆:第一轮PCR以5'RACE cDNA为模板。上游引物NUP,下游引物*psy*-5GSP-Ro,扩增程序:95℃预变性5 min,40个PCR循环(95℃变性30 s,68℃退火40 s,72℃延伸1 min),及72℃延伸5 min。PCR产物稀释50倍后作为第二轮PCR模板,下游引物替换为*psy*-5GSP-Ri,扩增程序:95℃预变性5 min,5个PCR循环1(95℃变性30 s,72℃退火延伸2 min),5个PCR循环2(95℃变性30 s,70℃退火30 s,72℃延伸2 min),30个PCR循环3(95℃变性30 s,68℃退火30 s,72℃延伸2 min),及72℃延伸5 min。

所有PCR产物经1%琼脂糖凝胶电泳后切胶回收,经普通琼脂糖凝胶DNA回收试剂盒(天根)纯化。纯化产物连接至pMD19-T载体(TaKaRa),并转化至感受态细胞DH5a(天根)。挑取阳性单克隆经PCR验证后送至上海生工生

物工程有限公司测序。

表 1 引物序列及 T_m 值

Tab.1 Sequence and annealing temperature of the primer

引物 Primers	序列 5'-3' Sequences 5'-3'	T_m /°C Annealing temperature
psy-pf	ACSAGCGARTATGCCAAGAC	53.6
psy-pr	CKCGCAGRATGTTKGTGTCAGC	50.9
psy-yf	GCCATCTACGTGTGGTGCC	66.1
psy-yr	CACGCAAGATGTGGTCAGC	65.1
psy-3GSP-Fo	CGCCACGTATGACGAGCTGTACGACTA	66.1
psy-3GSP-Fi	ACAAGGGGCCTCTGAGAAAGGTGTACC	70.7
psy-5GSP-Ro	CTTACCAGGTCCATGCGCATCCCTCTCC	72.5
psy-5GSP-Ri	CCTCAAAGATGCGCTCCAGCCGCTCTCTC	72.5
NUP	AAGCAGTGATCAACGCAGACT	63.9

1.4 序列分析

使用 DNAMAN (8.0, Lynnon) 将上述 *psy* 基因保守片段、3' 及 5' 末端片段序列拼接后获得 cDNA 全长序列。使用以下工具预测分析基因开放阅读框(ORF)、蛋白质序列、基本理化性质、功能位点、亚细胞定位,以及多序列比对、构建发生树,并构建蛋白质高级结构的三维模型。

Translate (<http://web.expasy.org/translate/>): 预测基因开放阅读框及氨基酸序列。

ProtParamon (<http://web.expasy.org/protaram>): 计算蛋白质的理论等电点和分子量。

BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>): 将序列与 GenBank 数据库进行对比。

MEGA 7.0: 计算遗传距离 (Poison 模型) 及构建系统发生树 (N-J 法, neighbor-joining), bootstrap 重复 1 000 次检验。

TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>): 预测蛋白质跨膜结构域。

TargetP (<http://www.cbs.dtu.dk/services/TargetP/>): 分析产物蛋白的亚细胞定位^[16]。

MEME (<http://meme-suite.org/tools/meme>): 保守区块分析。

PROSITE (<http://prosite.expasy.org/>): 基序 (motif) 分析。

NCBI Conserved domains (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>): 基序及结构域分析。

Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>): 蛋白质三维结构建模。

Swiss-PdbViewer (4.1, DeepView): 模型处理

及标记。

2 结果

2.1 *psy* 基因 cDNA 全长的克隆

对已报道的 *psy* 基因序列进行比对后,根据保守序列设计了两对引物(表 1),通过巢式 PCR 和降落式 PCR,扩增出了一段 406 bp 的片段。测序及 BLAST 数据库比对表明该片段序列与其他绿藻的 *psy* 基因有较高的相似性,推测该片段属于 *psy* 基因保守序列的一部分。根据这一片段序列,设计基因特异性引物(表 1)进行 RACE-PCR 扩增,分别获得了 5' 末端及 3' 末端序列。使用 DNAMAN 将上述 3 段序列拼接后,最终获得了完整的 *psy* cDNA 序列。

克隆得到的 cDNA 序列全长为 1 605 bp。使用 Translate (ExPASy) 对序列进行翻译并预测其开放阅读框,得到编码区序列长 1 248 bp,所编码的蛋白质包含 415 个氨基酸。编码区外分别具有 338 bp 5'-非编码区 (UTR, untranslated dregion) 及 22 bp 的 3'-UTR (图 1)。

经 ProtParam (ExPASy) 计算,预测的 PSY 蛋白质分子量为 46.83 ku,理论等电点为 8.92,含量最多的氨基酸为丙氨酸 (12.8%),不稳定系数 (instability index) 为 47.08,脂肪系数 (aliphatic index) 为 79.59,总平均亲水性 (GRAVY, grand average of hydropathicity) 为 -0.374。

通过 TMHMM 程序分析,该 PSY 蛋白不具有跨膜结构域,表明其可能不结合在膜上。

使用 TargetP 服务器分析所预测蛋白的亚细胞定位,结果显示在该 PSY 氨基酸序列中,1 ~ 45 位氨基酸残基可能是一段叶绿体转运肽 (cTP,

chloroplast transit peptide), 可信度为 3, 1 为最可信, 5 为不可信。

```

1 gaccctcgaattgttaaagtcgtctgtttcttgtgtctcctgacttcagtcogctgtctcagctgggtctctcgg
76 gacacaacagcacataaagggctgcccagcaagatttaggaagttgcagggagccctctgcgatataagcaacg
151 gagaattgttgcattgcacatatttcttcttcttggcggaccgacacccgactttctgtagaagcgtcagcg
226 cttcaccgggaatcagcaacaggggctcctcagaaacagcgtgtttcttgcggcaacttgagatctgaaaatccaat
1 M E A G L G S T Q P A A A
301 ctaactttgttgagaagaaaagtttcttggaaagcaatcATGGAGGCAGGCTTGGTTCTACCCAGCCCGCAGCTG
14 L G T A L P R R L Q F R P M T S R R R R P C K A L
376 CTCTGGGCACAGCTCTACCACGCCGGTGCAGTTCGGCCGATGACAAGCCGTAGGAGGGCCCTGCAAGGCGC
39 F V N N T A T V G N K Q A L A S D P D E P Y T S A
451 TTTTCGTGAACAACACAGCAACAGTCGGAACAAGCAGGCCGTGGCCCTCCGACCCCGACGAGCCTTACACCTCGG
64 P T K A P A A P R K L T G Q E I E E A A M W T Q I
526 CCCCCACTAAGGCACCTGCTGCTCCTCGAAAGCTTACAGGCCAGGAGATTGAGGAGGCAGCGATGTGGACGAGA
89 K L H E A Y P N G R S I P H N W T S K D L E A A Y
601 TCAAGCTGCACGAGGCATACCCCAACGGGAGGTCCATCCCCATAACTGGACCTCCAAAGACTTAGAGGCGCCCT
114 T R C A R V T S G Y A K T F Y L G T Q L M T P E K
676 ACACCAGGTGTGGCCGTGTACTAGTGGGTACGCCAAGACTTTCTATCTGGGGACTCAGTGTATGACACCCGAGA
139 A R A I W A I Y V W C R R T D E L V D G P N A A R
751 AAGCCCGAGCAATTTGGCCATCTACGTGTGGTGGCCGGACTGACGAGCTGGTGGACGGCCCTAACGCTGCCC
164 I T P A A L D R W E R L E A I F E G R P Y D S L
826 GCATCACACCTGCGGCTCTGGACCGGTGGGAGGAGCGGCTGGAGGCCATCTTTGAGGGTCGTCCTACGACAGCC
189 D A A L T D T I S R F P V D I Q P F R D M V E G M
901 TGGATCCGCCCCTCAGTACACCATCTCCCGCTTCCCGTGGACATCCAGCCCTTCCGGGACATGGTGGAGGGGA
214 R M D L V K S R Y A T Y D E L Y D Y C R V A G T
976 TGCCATGGACCTGGTGAAGTCGCGGTACGCCACGTATGACGAGCTGTACGACTACTGCTACAGGGTGGCGGGCA
239 V A L M S C P V M G V D P S Y K G P L E K V Y R A
1051 CGGTGACCTTGATGTCGTGCCCCGTCATGGGGTGGACCCCTCTACAAGGGGCCCTTGGAGAAAGGTGTACCGGG
264 A L A L G T A N Q L T N I L R D V G E D A S Q R N
1126 CGGCCCTGGCCCTAGGGACGGCAACCAGCTGACCAACATCCTGAGGGATGTGGGTGAGGACGCCCTCCAGCGCA
289 R I Y V P L E D L D R F S I K E E E V M R G M F A
1201 ACAGGATCTACGTCCCGTGGAGGACCTGGATAGGTTTCAGCATCAAGGAGGAGGAGGTGATGCGCGCATGTTG
314 P S T G R V D D R W R G F M A F Q I D R A R R I F
1276 CTTCCCTCCAGGGGCGCGTGGACGACCGGTGGCGGGTTTATGGCTTCCAGATAGACCGCGCCCGCGCATCT
339 A E A E A G V N L L D K D A R W P V W S A L I L Y
1351 TTGCTGAAGCGGAGGCGGGGTCAACCTGCTGGACAAGGACGCGCCAGGTGGCCGGTGTGGTTCGGCCCTGATTCTGT
364 R Q I L D G I E A N K Y D N F T K R A Y V P K F R
1426 ACCGCCAGATCTGGACGGCATCGAGGCCAACAAGTATGACAACCTTACCAAGCGCGCATATGTGCCCAAGITTA
389 K L A S L P L A L V A A R M P V N L V E Q H A S H
1501 GGAAGCTGGCGTCCCTGCCCCCTGGCTTGGTGGCAGCCCGCATGCCTGTCAATCTGGTGGAGCAGCATGCTAGCC
414 A L *
1576 ACGCATTGTGACatgtgaaatctcagccccgaaaaaaaaaaaaaaaaaaaaa

```

图 1 普通小球藻 *psy* 的 cDNA 全长序列及其预测的氨基酸序列

Fig. 1 The full-length cDNA and deduced amino acid sequences of *psy* in *C. vulgaris*

* 表示终止密码子

* indicates stop codon

2.2 同源性及系统发生树的构建

将普通小球藻的 PSY 氨基酸序列与 GenBank 数据库进行 Blastp 对比分析, 结果表明, 普通小球藻的 PSY 氨基酸序列与其他绿藻具有较高的相似性和一致性: 与原壳小球藻 (*Auxenochlorella protothecoides*)、小球藻 (*Chlorella variabilis*)、佐夫色绿藻 (*Chromochloris zofingiensis*)、莱茵衣藻 (*Chlamydomonas reinhardtii*)、绿色鞭毛藻 (*Ostreococcus lucimarinus*)、盐生杜氏藻 (*Dunaliella salina*) 的相似性分别为 84%、83%、81%、87%、85% 和 80%; 一致性分别为 76%、74%、69%、74%、73% 和 74%。与高等植物、蓝藻及细菌的相似性和一致性相对较低: 比如与百合 (*Lilium brownii*) 相似性

为 78%, 一致性为 67%; 与聚球蓝细菌 (*synechococcus* sp.) 相似性为 72%, 一致性为 58%; 与类球红细菌 (*Rhodobacter sphaeroides*) 相似性为 48%, 一致性为 32%。

将该 PSY 氨基酸序列与 48 种不同物种的 PSY 序列通过 Clustal Omega 进行比对, 并用 MEGA 7.0 构建系统发生树 (N-J 法)。从系统发生树 (图 2) 中可以看出, 不同物种的 PSY 蛋白质聚集成细菌、蓝细菌、绿藻和高等植物 4 个分支。预测所得的 PSY 氨基酸序列 (图 2 中 * 标示) 与其他绿藻聚集成一支, bootstrap 值为 92%。与普通小球藻在系统发生上最相近的是小球藻、原壳小球藻和佐夫色绿藻, 遗传距离分别为 0.173、0.188 和 0.239 (Poisson 模型)。此外, 高等植物

与绿藻的 PSY 蛋白聚集成一大类(bootstrap 值为 93%),表明其可能由同一个蛋白质进化而来。

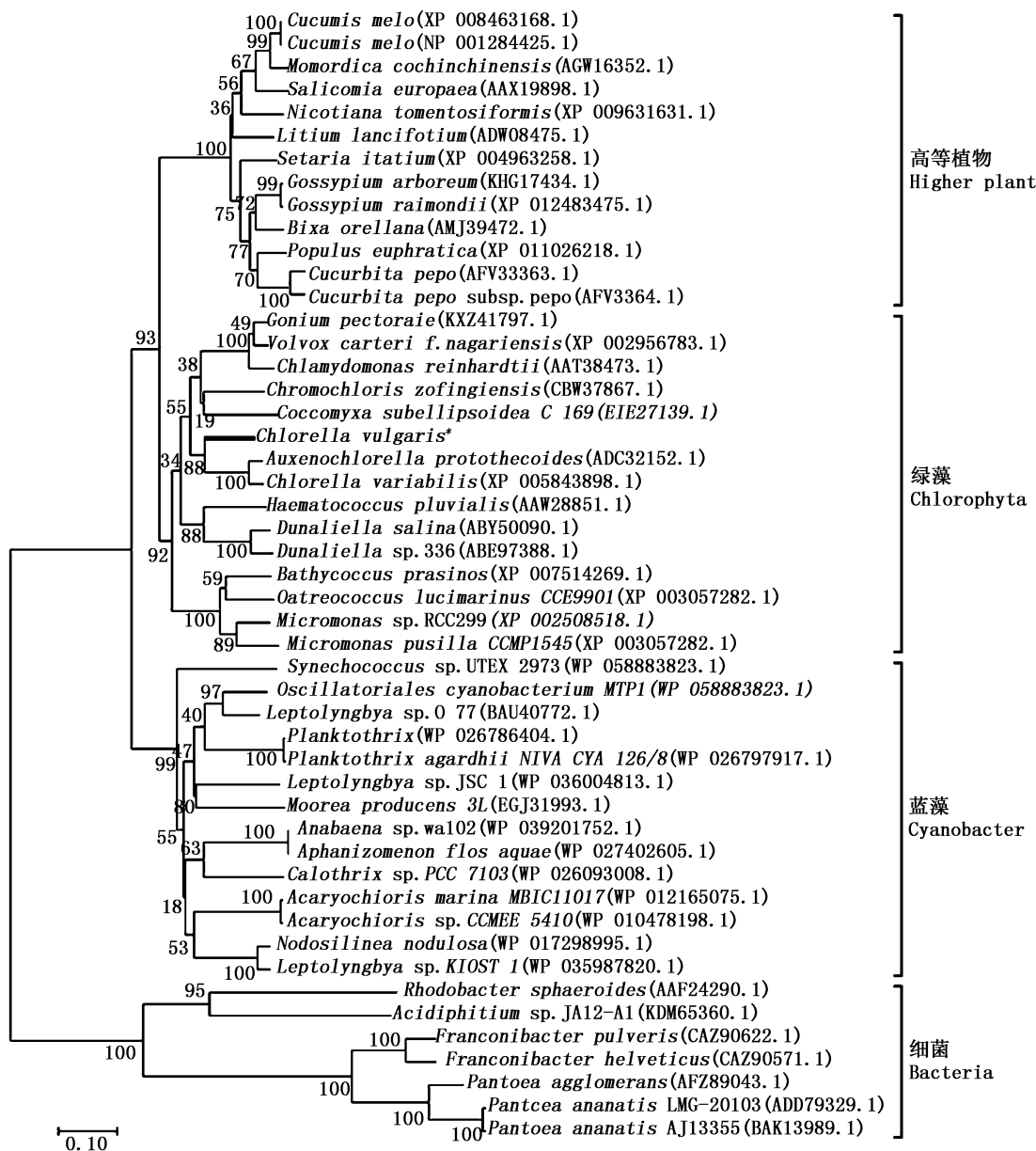


图 2 不同物种 PSY 氨基酸序列构建系统发生树(N-J 法)

Fig. 2 Phylogenetic tree of PSY amino acid sequences from various species (Neighbor-joining)

分支上的数字为 Bootstrap 值

Bootstrap values are shown at the branch points

2.3 保守区块 (Block) 及基序 (motif) 分析

将系统发生树(图 2)中的 48 种 PSY 氨基酸序列经由 MEME 服务器分析保守区块 (Block), 结果见图 3(图中仅列出部分物种)。所有序列共有 1 个保守区块,在普通小球藻的 PSY 氨基酸序列中位于 258 ~ 307 位,将其命名为 Block2。另有

2 个保守区块 Block1 (112 ~ 253 位) 和 Block3 (315 ~ 394 位) 只存在于高等植物、绿藻及蓝藻中(图 3a)。细菌的 *crtB* 序列(即 PSY)与其他物种的 PSY 蛋白质仅有一部分保守序列具有同源性 (Block2)。

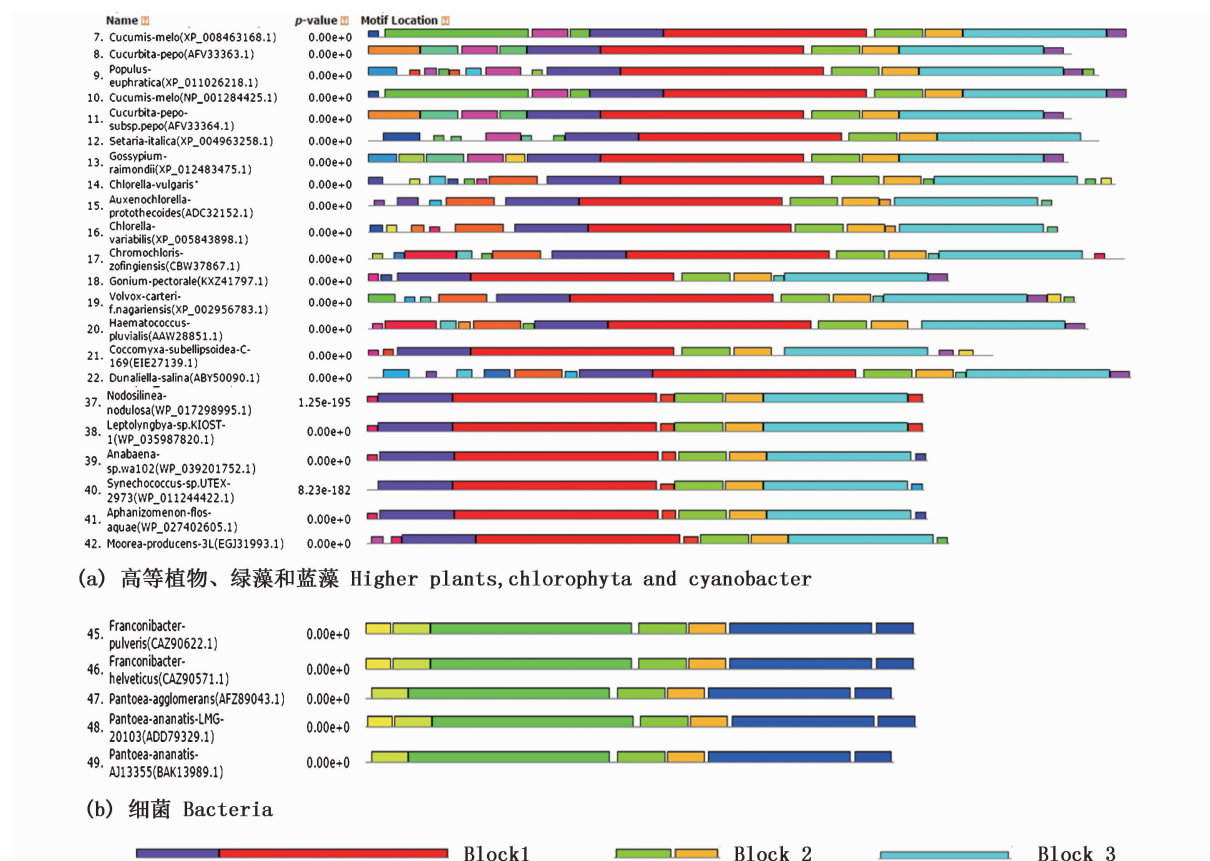


图3 部分物种 PSY 中的保守区块

Fig. 3 Conserved blocks in PSY from part organisms

通过 Prosite 和 NCBI Conserved domains 服务器对普通小球藻 PSY 氨基酸序列进行保守基序 (motif) 分析, 结果标注于图 4。在所有比对的 PSY 氨基酸序列中都存在 1 个保守基序, 在普通小球藻中位于 267 ~ 293 位氨基酸 (Block 2), 序列为 LGTANQLTNILRDVGEDASQRNRIYVP, 将其命名为 Motif S (图 4, 红色框标注)。数据库对比显示这段序列名为 “Squalene and phytoene synthases signature 2”, 正式表达式为 LGTanQlt. NIIRDVgeDasqrm. . RiYiP, 是八氢番茄红素合成酶蛋白的特征序列。在这段序列中存在一个保守的底物-镁离子结合位点/天冬氨酸富集区 (substrate-Mg²⁺ bindingsite/aspartate-rich region), 命名为 Motif D3, 起始氨基酸 279 位, 序列为 DVGED。另一个保守 DXXXD 基序在 Block1 中, 命名为 Motif D1, 起始氨基酸 153 位, 序列为 DELVD。在普通小球藻中存在第 3 个 DXXXD 基序, 位于 226 位氨基酸, 序列为 DELVD, 命名为

Motif D2 (图 4, 绿色框标注), 在所对比的 49 种 PSY 中, 只有普通小球藻、原壳小球藻和小球藻 3 种绿藻中存在 Motif D2 (图 4 仅列出部分绿藻)。此外, 在序列的第 123 位和第 381 位分别存在两个保守的活性位点遮蔽残基 (active site lid residues) YAKTF 和 RAYV, 分别命名为 Motif L1 和 Motif L2 (图 4, 蓝色框标注)。

除上述保守基序外, 序列中存在多个潜在的蛋白质修饰位点, 列于表 2。其中一部分位点位于保守区块中, 可能是 PSY 蛋白的特征修饰位点。有 3 个位点虽然位于保守区块中, 但却与其他物种的保守序列不同, 可能是普通小球藻特有的修饰位点。

在序列的 114 位氨基酸 (Block1) 之前, 还有多个非保守的潜在修饰位点, 包括 2 个天冬酰胺-糖基化位点、3 个 cAMP 和 cGMP 依赖的蛋白激酶磷酸化位点、3 个酪氨酸激酶磷酸化位点以及 3 个酰基化位点 (表 2, 图 4 中黑色框标注)。

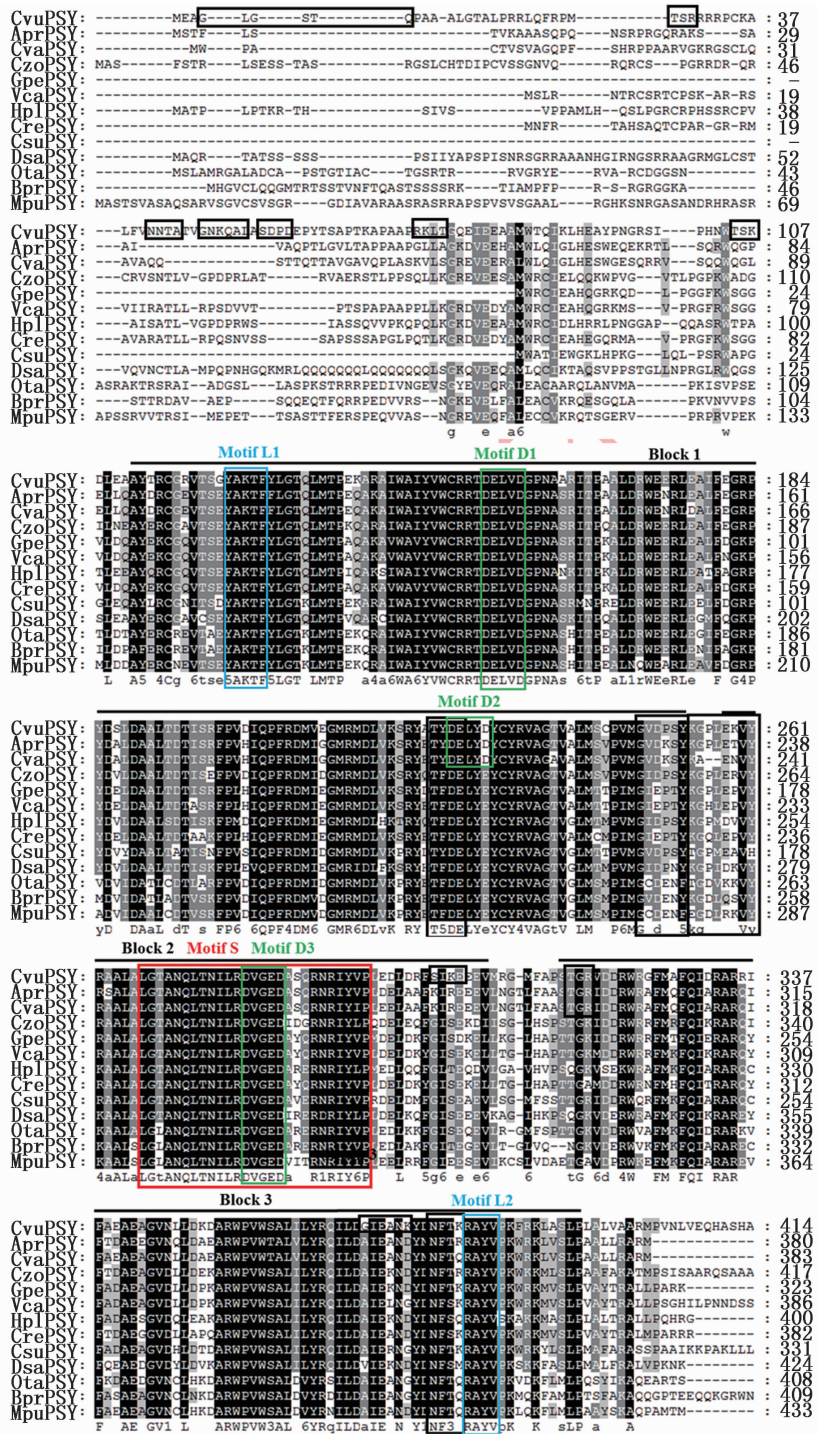


图 4 部分绿藻 PSY 氨基酸多序列对比

Fig. 4 Multiple sequences alignment of PSY amino acids in some chlorophyta species

1. 保守区块 (Block 1, 2, 3) 用横线标出, 红色框表示 PSY 特征序列 (Motif S), 绿色框表示底物-镁离子结合位点/天冬氨酸富集区 (Motif D1, D2, D3), 蓝色框表示活性位点保护基序 (Motif L1, L2), 部分蛋白质修饰位点用黑色框标出。Conserved blocks (Block 1, 2, 3) are under black lines. "Squalene and phytoene synthases signature 2" (Motif S) is in red box. Substrate-Mg²⁺ binding sites/aspartate-rich regions (Motif D1, D2, D3) are in green boxes. Active site lid residues (Motif L1, L2) are in blue boxes. Part of modified sites are in black boxes; 2. CvuPSY: 普通小球藻 *Chlorella vulgaris*; AprPSY: 原壳小球藻 *Auxenochlorella protochloides* (ADC32152.1); CvaPSY: 小球藻 *Chlorella variabilis* (XP_005843898.1); CzoPSY: 佐夫色绿藻 *Chromochloris zofingiensis* (CBW37867.1); GpePSY: 胸状盘藻 *Gonium pectorale* (KXZ41797.1); VcaPSY: 团藻 *Volvox carterif nagariensis* (XP_002956783.1); HplPSY: 雨生红球藻 *Haematococcus pluvialis* (AAW28851.1); CrePSY: 莱茵衣藻 *Chlamydomonas reinhardtii* (XP_001701192.1); CsuPSY: 胶球藻 *Coccomyxa subellipsoidea* C-169 (EIF27139.1); DsaPSY: 盐生杜氏藻 *Dunaliella salina* (ABY50090.1); OtaPSY: 鞭毛藻 *Ostreococcus tauri* (XP_003079460.1); BprPSY: 海洋青绿藻 *Bathycoccus prasinos* (XP_007514269.1); MpuPSY: 细小微胞藻 *Micromonas pusilla* CCMP1545 (XP_003057282.1)

表 2 普通小球藻 PSY 氨基酸序列中的修饰位点
Tab. 2 Modification sites in the PSY amino acid sequence of *C. vulgaris*

名称 Name	起始氨基酸位(保守区) Position of the first amino acid (Block)	序列 Sequence
天冬酰胺-糖基化位点 N-glycosylation site	41	NNTA
	103	NWTS
	377 (Block 3)	NFTK
cAMP 和 cGMP 依赖的蛋白激酶磷酸化位点 cAMP- and cGMP-dependent protein kinase phosphorylation site	72	RKLT
蛋白激酶 C 磷酸化位点 Protein kinase C phosphorylation site	28	TSR
	105	TSK
	252 (Block 1)	SYK
	285 (Block 2)	SQR
	301 (Block 2) *	SIK
	316 (Block 3) *	TGR
	379 (Block 3)	TKR
酪蛋白激酶 II 磷酸化位点 Casein kinase II phosphorylation site	54	SDPD
	75	TGQE
	105	TSKD
	224 (Block 1)	TYDE
	301 (Block 2) *	SIKE
酪氨酸激酶磷酸化位点 Tyrosine kinase phosphorylation site	107	KDLEAAY
	254 (Block 2)	KGPLEKVY
酰基化位点 N-myristoylation site	4	GLGSTQ
	47	GNKQAL
	248 (Block 1)	GVDPSY
	369 (Block 3) *	GIEANK

注: * 号表示保守区块中的非保守位点

Note: * indicates nonconservative sites in conservative block

2.4 PSY 蛋白质的高级结构建模

通过 Phyre2 服务器,使用同源建模法(homology-based)与从头计算法(*ab initio*)预测 PSY 蛋白质的二级结构和三级结构三维模型。首先通过 PDB 数据库(Protein Data Bank)对比,依据序列一致性、比对的覆盖率及模板的可信度,以启发式算法(heuristics)选出了 4 个综合最优的模板(c2zcpA, c4hd1A, c3we9A, d1ezfa),这些模板为细菌中已知的八氢番茄红素或角鲨烯(squalene)合成酶的晶体结构。根据这些模板,首先将 PSY 蛋白序列中与之同源的部分建模,这段序列为 104 ~ 405 位氨基酸残基,这部分模型的可信度为 90% 以上(Phyre2 服务器评估)。剩余的 1 ~ 103 位及 406 ~ 415 位残基使用从头计算法建模,可信度 < 90%。最终将三段模型拼接后形成 PSY 蛋白质的三维模型。

预测得到的二级结构如图 5,序列中共有 21 个 α 螺旋,65 个 β 转角以及 12 个 γ 转角。预测构建的 PSY 蛋白三维模型见图 6,图中标出了蛋

白质肽链的氨基端(N 端)和羧基端(C 端)。4 个保守基序(Motif D1、D3, Motif L1、L2)和非保守 Motif D2 的位点及结构在图中用白色表示。

3 讨论

在食品及医药等领域中,类胡萝卜素所蕴含的实用价值使得人们致力于通过生物合成,特别是从藻类中获取更多类胡萝卜素产物^[8]。在类胡萝卜素生物合成途径中,八氢番茄红素合成酶 PSY 是催化关键步骤的限速酶,影响后续类胡萝卜素合成的反应速率^[7]。除了通过特定培养条件增加藻细胞内类胡萝卜素的累积外,已有报道在植物和微藻中通过基因修饰构建外源的 *psy* 过表达载体,成功提高了宿主细胞内类胡萝卜素的含量^[10, 17]。普通小球藻易于培养,常作为藻类研究的模式生物且富含类胡萝卜素,对于普通小球藻类胡萝卜素相关基因的研究还较少,本研究首次对普通小球藻 *psy* 基因进行了克隆(图 1)。

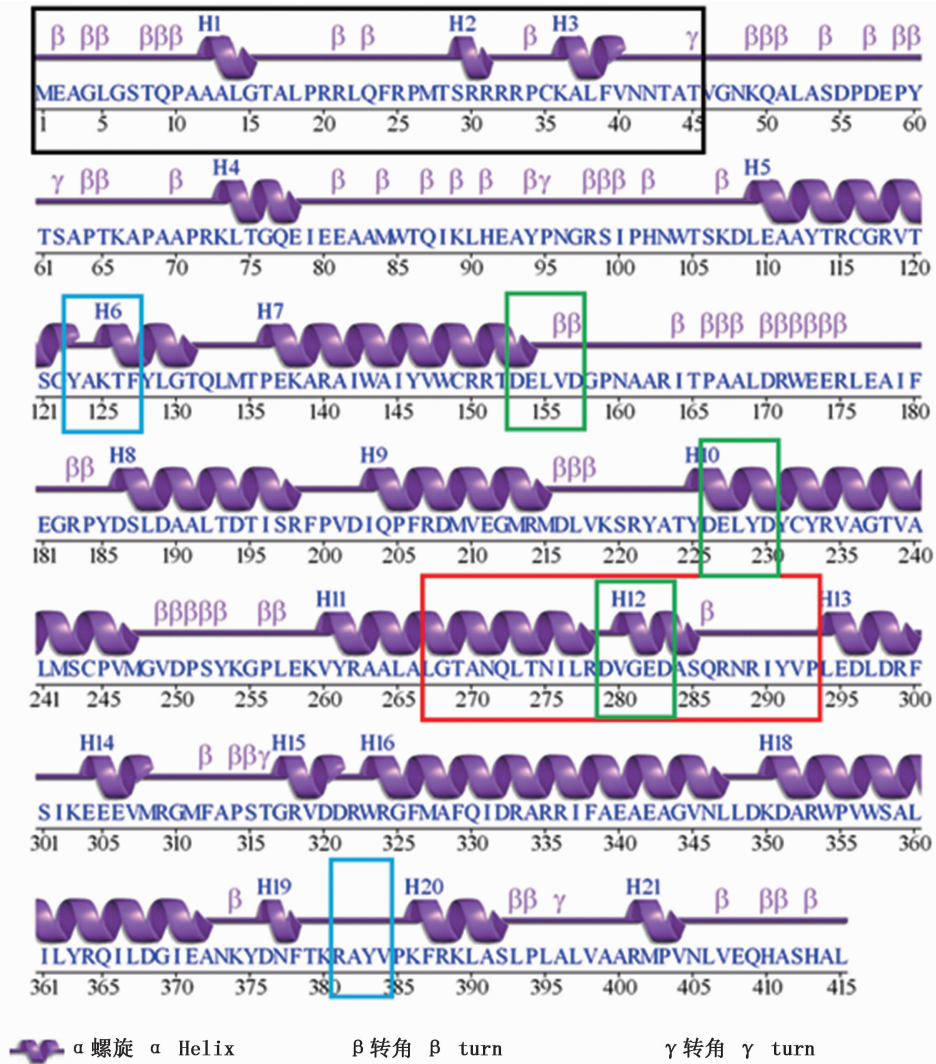


图 5 普通小球藻 PSY 蛋白质序列的二级结构

Fig. 5 The secondary structure of PSY protein sequence of *C. vulgaris*

黑色框为预测的叶绿体转运肽,红色、绿色和蓝色框为保守基序(参照图 4)

Deduced chloroplast transit peptide is in black box. Motifs are in red, green and blue boxes (refer to fig. 4)

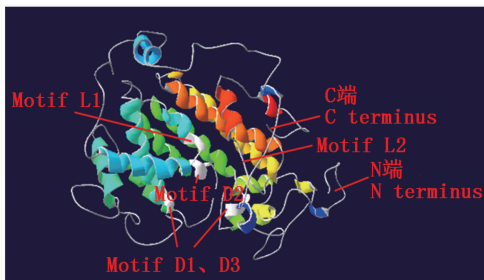


图 6 预测的普通小球藻 PSY 蛋白质高级结构

Fig. 6 Putative tertiary structure of PSY protein in *C. vulgaris*

将克隆得到的 cDNA 序列翻译成氨基酸序列后,通过多序列比对发现,该序列与其他绿藻中

的 PSY 蛋白具有较高的相似性和一致性。系统发生分析也表明,该序列与绿藻的 PSY 蛋白具有同源性,与小球藻、原壳小球藻和佐夫色球藻的遗传距离较为接近(图 2)。对序列中的保守区块及功能基序进行分析,发现序列中存在 PSY 蛋白的保守区块(图 3),在这些区块中包含了 PSY 蛋白的特征序列“Squalene and phytoene synthases signature 2”(Motif S),该特征序列可能与底物结合及催化活性有关^[18]。此外还有 2 个“底物-镁离子结合位点”(Motif D1、D3)和 2 个“活性位点遮蔽残基”(Motif L1、L2),均属于 PSY 蛋白序列的保守基序。以上结果从同源性、系统发生及保守基序 3 方面可以推断在普通小球藻中克隆获

得的全长 cDNA 序列为八氢番茄红素合成酶的编码基因 *psy*。

目前已经获得了普通小球藻叶绿体的全基因组序列 (GenBank 登录号: NC_001865.1)^[19], 其中包括编码光反应中心蛋白、叶绿素合成酶、叶绿体分裂相关的基因。使用 Blastn 将 *psy* 的 cDNA 序列与叶绿体基因组序列进行比对后, 并未在基因组序列中找到 *psy* 基因, 表明 *psy* 不属于叶绿体基因组。由于类胡萝卜素的合成途径位于叶绿体中^[20], 因此 PSY 蛋白质可能在叶绿体外翻译后进入叶绿体内。亚细胞定位 (TargetP) 的分析结果支持了这一推测, PSY 序列的前 45 个氨基酸可能为一段叶绿体转运肽 (图 5)。转运肽能引导肽链与叶绿体外被膜结合, 转运肽穿过叶绿体膜后被切除, 剩余肽链折叠修饰后形成完整蛋白质进入叶绿体^[21-22], 相应地, PSY 序列上具有多个潜在的蛋白质修饰位点 (表 2), 可用于蛋白质翻译后的修饰以完成折叠。

使用同源建模和从头计算相结合的方法, 构建了普通小球藻 PSY 蛋白质的三维结构模型 (图 6)。在该模型中, 两个保守的“底物-镁离子结合位点” Motif D1、D3 分别位于相对的两个壁上, 可各结合一个底物二牻牛儿基二磷酸 (GGPP)。随后的催化反应分成两步, 第一步催化两分子 GGPP 生成 1 分子前植物烯二磷酸 (prephytoenediphosphate) 和 1 分子焦磷酸, 第二步由前植物烯二磷酸脱去二磷酸后生成八氢番茄红素 (phytoene)^[23-24]。由于是两步反应, 因此在底物结合位点附近存在两个“活性位点遮蔽残基” Motif L1 和 L2, 能够从细胞环境中保护高活性的中间产物^[25]。

早先报道在原壳小球藻和小球藻两种绿藻的 PSY 蛋白序列中存在第 3 个“底物-镁离子结合位点”^[11], 在普通小球藻的 PSY 预测氨基酸序列中同样发现了这一基序 Motif D2 (序列 DELYD), 而在其他所对比物种中该序列主要为 DELYE, 并不具有“底物-镁离子结合位点”的特征。在预测的三维模型中, Motif D2 的二级结构主要为 α 螺旋, 位于 Motif D1 和 Motif D3 的附近, 三者组成环形结构, 与原壳小球藻的结构相似^[11]。该位点是否具有活性, 在功能上对 PSY 蛋白有何影响还未知, 后续可对不同 PSY 蛋白质进行比较, 或应用定点突变^[26] 方法对其进行研

究。

综上, 本文首次克隆了普通小球藻中八氢番茄红素合成酶基因 *psy* 的 cDNA 全长, 预测其蛋白质序列, 构建了蛋白质的高级结构三维模型, 并分析了其底物结合位点及催化反应过程。所获得的结果可为相关基因及蛋白质研究提供基因序列及理论预测, 同时为构建过表达载体提供基因信息, 从而有利于提高类胡萝卜素产量, 满足日益上升的市场需求。

参考文献:

- [1] SUN Z, LI T, ZHOU Z G, et al. Microalgae as a source of lutein: chemistry, biosynthesis, and carotenogenesis [M]// POSTEN C, CHEN S F. Microalgae Biotechnology. Switzerland: Springer, 2016: 37-58.
- [2] BAI W S, WANG X W, WANG Z C, et al. The effect of astaxanthin (AST) on Neurotrophin-3 (NT-3) expression in rats after compressive spinal cord injury (SCI) [J]. African Journal of Pharmacy and Pharmacology, 2012, 6 (34): 2559-2654.
- [3] LI S Y, FUNG F K C, FU Z J, et al. Anti-inflammatory effects of lutein in retinal ischemic/hypoxic injury: *in vivo* and *in vitro* studies [J]. Investigative Ophthalmology & Visual Science, 2012, 53(10): 5976-5984.
- [4] KOWLURU R A, MENON B, GIERHART D L. Beneficial effect of zeaxanthin on retinal metabolic abnormalities in diabetic rats [J]. Investigative Ophthalmology & Visual Science, 2008, 49(4): 1645-1651.
- [5] AGARWAL M, PARAMESWARI R P, VASANTHI H R, et al. Dynamic action of carotenoids in cardioprotection and maintenance of cardiac health [J]. Molecules, 2012, 17 (4): 4755-4769.
- [6] MOEHS C P, TIAN L, OSTERYOUNG K W, et al. Analysis of carotenoid biosynthetic gene expression during marigold petal development [J]. Plant Molecular Biology, 2001, 45 (3): 281-293.
- [7] SALVINI M, BERNINI A, FAMBRINI M, et al. cDNA cloning and expression of the phytoene synthase gene in sunflower [J]. Journal of Plant Physiology, 2005, 162(4): 479-484.
- [8] TAKAICHI S. Carotenoids in algae: distributions, biosyntheses and functions [J]. Marine Drugs, 2011, 9(6): 1101-1118.
- [9] LIANG C W, ZHAO F Q, QIN S, et al. Molecular cloning and characterization of phytoene synthase gene from a unicellular green alga *Haematococcus pluvialis* [J]. Progress in Biochemistry and Biophysics, 2006, 33(9): 854-860.
- [10] CORDERO B F, COUSO I, LEÓN R, et al. Enhancement of carotenoids biosynthesis in *Chlamydomonas reinhardtii* by nuclear transformation using a phytoene synthase gene

- isolated from *Chlorella zofingiensis*[J]. Applied Microbiology and Biotechnology, 2011, 91(2): 341-351.
- [11] LI M Y, CUI Y, GAN Z B, et al. Isolation and analysis of the *cpsy* gene and promoter from *Chlorella protothecoides* CS-41[J]. Marine Drugs, 2015, 13(11): 6620-6635.
- [12] SUN Z, PENG X F, LIU J, et al. Inhibitory effects of microalgal extracts on the formation of advanced glycation endproducts (AGEs)[J]. Food Chemistry, 2010, 120(1): 261-267.
- [13] STARA A, SERGEJEVOVA M, KOZAK P, et al. Resistance of common carp (*Cyprinus carpio* L.) to oxidative stress after chloramine-T treatment is increased by microalgae carotenoid-rich diet[J]. Neuroendocrinology Letters, 2014, 35(s2): 71-80.
- [14] SAFI C, ZEBIB B, MERAH O, et al. Morphology, composition, production, processing and applications of *Chlorella vulgaris*: a review[J]. Renewable and Sustainable Energy Reviews, 2014, 35: 265-278.
- [15] RIPPKA R, DERUELLES J, WATERBURY J B, et al. Generic assignments, strain histories and properties of pure cultures of cyanobacteria[J]. Microbiology, 1979, 111(1): 1-61.
- [16] EMANUELSSON O, BRUNAK S, VON HEIJNE G, et al. Locating proteins in the cell using TargetP, SignalP and related tools[J]. Nature Protocols, 2007, 2(4): 953-971.
- [17] LINDGREN L O, STÅLBERG K G, HÖGLUND A S. Seed-specific overexpression of an endogenous arabidopsis phytoene synthase gene results in delayed germination and increased levels of carotenoids, chlorophyll, and abscisic acid[J]. Plant Physiology, 2003, 132(2): 779-785.
- [18] SUMMERS C, KARST F, CHARLES A D. Cloning, expression and characterisation of the cDNA encoding human hepatic squalene synthase, and its relationship to phytoene synthase[J]. Gene, 1993, 136(1/2): 185-192.
- [19] WAKASUGI T, NAGAI T, KAPOOR M, et al. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division[J]. Proceedings of the National Academy of Sciences of the United States of America, 1997, 94(11): 5967-5972.
- [20] ROHMER M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants[J]. Natural Product Reports, 1999, 16(5): 565-574.
- [21] 吴光耀. 核编码叶绿体蛋白中的转运肽[J]. 植物生理学通讯, 1991, 27(3): 166-172.
- WU G Y. Transit peptides for nuclear encoded proteins import into chloroplasts [J]. Plant Physiology Communications, 1991, 27(3): 166-172.
- [22] 赵海铭, 宋伟彬, 赖锦盛. 高粱 5-烯醇丙酮酰莽草酸-3-磷酸合酶基因(EPSPS)叶绿体转运肽(CTP)的克隆及其在转基因玉米中的功能验证[J]. 农业生物技术学报, 2013, 21(9): 1009-1018.
- ZHAO H M, SONG W B, LAI J S. Cloning of Sorghum bicolor chloroplast transit peptide (CTP) of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) and its functional validation in transgenic maize (*Zea mays*) [J]. Journal of Agricultural Biotechnology, 2013, 21(9): 1009-1018.
- [23] LADYGIN V G. Biosynthesis of carotenoids in the chloroplasts of algae and higher plants [J]. Russian Journal of Plant Physiology, 2000, 47(6): 796-814.
- [24] LI F Q, VALLABHANENI R, YU J, et al. The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress tolerance [J]. Plant Physiology, 2008, 147(3): 1334-1346.
- [25] LIANG P H, KO T P, WANG A H J. Structure, mechanism and function of prenyl transferases[J]. European Journal of Biochemistry, 2002, 269(14): 3339-3354.
- [26] 郭子平, 翟清华, 曾令锋, 等. 小麦 *TaGAPC* 定点突变体基因载体构建及其原核表达[J]. 华北农学报, 2016, 31(1): 46-50.
- GUO Z P, ZHAI Q H, ZENG L F, et al. Construction of *TaGAPC* site-directed mutagenesis gene vector and prokaryotic expression analysis from wheat [J]. Acta Agriculturae Boreali-Sinica, 2016, 31(1): 46-50.

cDNA cloning and bioinformatics analysis of the phytoene synthase encoding gene *psy* in *Chlorella vulgaris*

LIU Dongyu¹, TAO Xianji¹, WEI Hua²

(1. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China; 2. Shanghai Vocational College of Agriculture and Forestry, Shanghai 201699, China)

Abstract: The full-length cDNA sequence of phytoene synthase (PSY) encoding gene *psy* in *Chlorella vulgaris* (*C. vulgaris*) was isolated by RACE-PCR. The 1605 bp sequence contained a 1245 bp open reading frame (ORF), which encoded a protein composed of 415 amino acids residues. Results of multiple sequence alignment and phylogeny showed that the deduced protein sequence of the *C. vulgaris* PSY was homologous with other PSY proteins, and clustered into the green algae group. The genetic distances of PSY between *C. vulgaris* and *Chlorella variabilis*, *Auxenochlorella protothecoides*, *Chromochloris zofingiensis* was 0.173, 0.188 and 0.239 respectively. And those PSYs also kept the high similarity (81% -84%) and identity (69% -76%). The prediction result of subcellular location showed that the amino acids residues from 1st to 45th might perform as a chloroplast transit peptide, suggesting that PSY must be translated outside before entering the chloroplast and being active. Analyses of conserved blocks/motifs and tertiary structure modelling indicated that this PSY sequence contained several potential modification sites, with the conserved blocks and characteristic motifs of PSY protein, including two substrate-Mg²⁺ binding sites and two active site lid residues which might mediate binding of substrate and shield highly reactive intermediates from solvent. A specific substrate-Mg²⁺ binding site (the third) was found in the *C. vulgaris* PSY, and its activity and function was still unconfirmed. Taken together, these results uncovered the full-length cDNA sequence of *psy*, and predicted the putative structure of PSY protein in *C. vulgaris*, and thus provided information for over-expression of *psy* and for enhancing the productivity of carotenoids.

Key words: phytoene synthase; *psy*; *Chlorella vulgaris*; carotenoids