In Silico Cloning and Sequence Analysis of F3H Gene in *Phaseolus coccineus*

MA Guang, GUO Ji-ping

(Department of Life Sciences, Hengshui College, Hengshui 053000, China)

Abstract: Flavanone 3-hydroxylase is a key enzyme in the biosynthetic pathway of plant flavonoids. The putative cDNA sequence of *Phaseolus coccineus* F3H gene was in silico cloned by using *Glycine max* F3H gene cDNA sequence (AAU06216.1) as a probe. According to the putative cDNA, the primers were designed for RT-PCR. The coding sequence of F3H gene in *Phaseolus coccineus* cDNA was cloned using RT-PCR. Then the secondary structure and advanced structure of F3H in *Phaseolus coccineus* were analyzed by bioinformatics methods. The results showed that the cDNA contains an open reading frame of 1128 bp, which encoded the protein containing 375 amino acids. The protein coded by *Phaseolus coccineus* F3H gene cDNA showed 92% similarity to *Glycine max*. Its secondary structure contained 40.8% α-helix, 14.93% extended strand, 39.73% random coil and 4.53% β-turn. The homology analysis of three-dimensional structure showed that the three-dimensional structure of the protein was a compact globular structure.

Key words: flavanone 3-hydroxylase; three-dimensional structure; sequence analysis; RT-PCR; bioinformatics

Article No.: 1673-9078(2014)4-6-9

红花菜豆黄烷酮-3-羟化酶基因电子克隆及序列分析

马光，郭继平

（衡水学院生命科学系，河北衡水 053000）

摘要：黄烷酮-3-羟化酶（F3H）是植物类黄酮物质生物合成途径中的一个关键酶。本文利用电子克隆的方法，采用已经报道的大豆 F3H 基因 cDNA 序列为种子序列，搜索红花菜豆 EST 数据库，获得了假定的红花菜豆 F3H 基因 cDNA 的序列。根据假定的 cDNA 序列设计引物，采用 RT-PCR 的方法获得了红花菜豆 F3H 基因 cDNA 的编码序列。并采用生物信息工具对红花菜豆 F3H 的性质包括氨基酸序列组成、物理和化学性质、二级和三级结构的特点进行了研究。结果表明，红花菜豆 F3H 基因 cDNA 包含一个 1128 bp 的完整的开放读码框，其编码蛋白由 375 个氨基酸组成。与大豆黄烷酮-3-羟化酶同源性为 92%。其二级结构含有 40.8% α 螺旋和 39.73% 无规则卷曲，β 折叠较少为 4.53%。采用同源模建的方式分析其三维立体结构，表明其三维结构是一个紧密的球状结构。

关键词：黄烷酮-3-羟化酶；三维结构；序列分析；RT-PCR；生物信息学

1 Introduction

Flavanone 3-hydroxylase (F3H; EC 1.14.11.9) is a key enzyme of flavonoid in plants. It catalyzes hydroxylation of C-ring third position hydroxyl catalytic at flavanone to form two hydrogen flavanols. It was firstly found in the crude extracts of violet, and its properties were studied by parsley cell culture [1]. Subsequently, F3H gene was cloned from the petunia and it showed high activity by functional expression in *Escherichia coli* [2]. Later, more characteristics of F3H gene in different species were reported, such as barley, begonia, alfalfa, maize, thale Cress and perilla [3]. However, the expression characteristics of F3H are different in diverse species. In petunia and snapdragon, regulation of flavonoid biosynthesis pathway of the preliminary step and late stage is different. For example, F3H gene belongs to the early gene in petunia, but belongs to the late genes in antirrhinum. In maize, the flavonoid biosynthesis pathway is through co-regulation to produce anthocyanin in organizations, but flavonol F3H gene expression is only consistent with accumulation of flavonol in anthers. In the study of alfalfa, F3H is also expressed in the roots and nodules, but its role is not clear [4].

Flavonoids have very high application value in the

收稿日期：2013-09-12

基金项目：河北省高等学校自然科学研究计划项目（Z2012045）；河北省科学技术研究与发展计划项目（12212702）

作者简介：马光（1977-），男，博士，副教授，研究方向功能性食品基因工程

medical and health care, *Leguminosae* and *Compositae* plants are rich in flavonoids [5]. As an important flavonoid biosynthesis gene, F3H gene of *Glycine* max had been cloned, however, it was reported little in other *Leguminous* plants, and no report was observed in *Phaseolus coccineus*. In this study, with the *Glycine* max F3H gene cDNA as probe, the F3H gene of *Phaseolus coccineus* was in silico cloned. The F3H cDNA coding sequence was obtained, the sequence and coding protein characteristics were investigated, and the three dimensional structure was constructed by homology model.

2 Materials and Methods

2.1 BLAST Searching of *Phaseolus coccineus* EST Databases

The cDNA sequence of *Glycine max* F3H gene (GenBank, AAU06216.1) was used as a probe to search the *Phaseolus coccineus* expressed sequence tag (EST) database through the BLAST program for a homologous clone. The EST sequence of score ≥100 and length ≥100 bps selected from the blast result, were generated contigs. The longer contig was used as second probe. The above step was not repeated until the newly generated probe could not be elongated. This approach led to a sequence as a putative *Phaseolus coccineus* F3H gene cDNA.

2.2 cDNA cloning of F3H gene from *Phaseolus coccineus* by RT-PCR

With 1383 bp cDNA sequence in silico cloned, one pair of RT-PCR primers were designed using Primer Premier 5.0 software. The primers were F3H-F (5'-TCACCATCATGGCTCCCACA-3') and F3H-R (5'-CCTCTGCTCAAATAAGGTGGT-3'), which were synthesized by Sangon Biotech (Shanghai) Co., Ltd. The primers corresponding to the RT-PCR with length of 19~1313 bp contained electronic cloning results. Total RNA was extracted from total *Phaseolus coccineus* seedling with age of 15 days using the method of CTAB. RT-PCR was performed with one cycle of 94 ℃ for 2 min followed by 30 cycles of denaturation (15 s at 94 ℃), anneal (30 s at 58 ℃), and extension (90 s at 62 ℃) and a final step of 72 ℃for 5 min. RT-PCR kit was ReverTra - Plus - (Code No. PCR-501) produced by Toyobo (Dalian, China). The PCR products were cloned into pGEM -T recovery easy vector (Promga Company) and were sequenced by Sangon Biotech (Shanghai) Co., Ltd.

2.3 Bioinformatics analysis of the *Phaseolus coccineus* F3H gene

The sequenced *Phaseolus coccineus* F3H gene cDNA was analyzed by bioinformatics software. The cDNA sequence was submitted to open reading frame software, and to find the start codon and termination codon position. The protein sequence of F3H gene of *Phaseolus coccineus*, *Glycine max*, *Medicago truncatula*, *Ipomoea batatas* and *Arabidopsis thaliana* were alignmented by multiple sequences alignment software Clustal Omega to compare their protein similarity. Primary structure characteristics of *Phaseolus coccineus* F3H protein and its secondary structure were analyzed using online server. The Neighbor-Joining phylogenetic tree was constructed by Mega 5.05 software.

2.4 The prediction of F3H protein three-dimensional structure

The three-dimensional structure of the protein was predicted using Swiss-Model online server, and showed by CN3D software.

3 Results

3.1 In silico cloning of *Phaseolus coccineus* F3H gene

With the cds of *Glycine max* F3H mRNA (GenBank, AAU06216.1) as probe, 10 EST sequences (score ≥100 and length ≥100) were found by Blast searching of the *Phaseolus coccineus* EST database, in NCBI.

The sequences were selected and saved in a file with FASTA format. The file was submitted to on-line CAP software. One contig was obtained by assembling EST sequences. It was putative *Phaseolus coccineus* F3H cDNA.

3.2 Verification of putative *Phaseolus coccineus* F3H cDNA by RT-PCR

With the total cDNA of *Phaseolus coccineus* seedling as template, one specific sequence was amplified by RT-PCR. The length of PCR products was about 1300 bp through analysis of agarose gel electrophoresis (Fig.1), which was consistent with what expected from in silico cloning. The sequencing result was similar to putative *Phaseolus coccineus* F3H cDNA. The results showed *Phaseolus coccineus* F3H cDNA was

cloned. Through searching the cDNA sequence for potential coding regions by ORF finder, an entire open reading frame (ORF) of 375 amino acids was detected (Fig.2 and Fig.3).

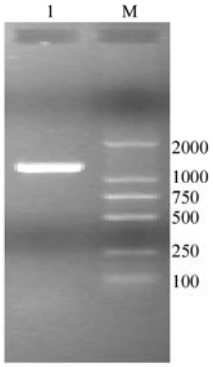

Fig.1 RT-PCR amplification of F3H cDNA from *Phaseolus coccineus* seedling

Note: 1: RT-PCR product; M: DNA marker (DL2000).

CACATCAAACTCATTTCACCATCATGGCTCCCACAGCCAAGACCCTCACGTAC
CTCGCGCAGGAGAAGACCCTGGAATCCAGCTTCGTCAGGGACGAGGACGAG
CGCCCCAAGGTTGCCTACAACCAGTTCAGCGACGAGATCCCCGTTATTTCTC
TCGCCGGAATCGACGACGTCCATGGCCGCAGACAGAGCATTTGCCAGAAAAT
AGTCCAGGCTTGCGAGGATTGGGGTATTTTTCAGGTTGTTGATCACGGCGTC
GATCAGAAACTCGTCTCCGAAATGACTCGTCTCGCCACCCAGTTCTTCGCTT
TGCCACCTGAGGAGAAGCTTCGCTTCGACATGTCCGGTGGCAAAAAGGGTG
GCTTCATTGTCTCCAGCCATCTCCAAGGAGAGTCGGTGCAGGACTGGAGAGA
GATTGTGACATACTTTTCGTACCCGAAGAAAGAAAGGGACTATTCGCGGTGG
CCGGACACACCAGAAGGGTGGAGAGCGGTGACTGAGGAATACAGTGAGAAG
CTGATGGGGCTGGCATGCAAATTGTTGGAAGTGTTGTCTGAGGCAATGGGGT
TAGAGAAGGAGGCTTTGACCAATGCATGTGTTGATATGGACCAGAAGGTTGT
GGTGAACTACTACCCCAAATGCCCTCAACCTGACCTCACACTCGGTCTCAAA
CGCCACACTGATCCTGGCACCATCACCTTGCTGCTTCAGGACCAAGTCGGAG
GACTTCAGGCCACCAGAGACAATGGTAAAACCTGGATCACTGTTCAGCCTGT
GGAGGGTGCTTTTGTTGTCAATCTTGGCGATCATGGTCACTTTGTGAGCAAT
GGAAGGTTCAAGAATGCTGATCACCAAGCAGTGGTGAACTCGAACCACAGCC
GTTTGTCGATAGCCACGTTTCAGAACCCAGCACCAAATGCAACTGTGTACCC
TTTGAAGGTGAGGGAGGGAGAGAAGCCTGTGCTGGAGGAACCCATCACTTT
TGCTGAAATGTACAAGAAGAAAATGAGCAAGGATCTTGAGATTGCAAGGATG
AAGAAGCTGGCTAAGGAAAAGCAATTGCAGGACCTTGAGAATGAAAGCATT
TGCAAGAACTTGATCAGAAGGCAAAGTTGGAGACCAAGCCTTTGAAGCAAAT
TCTTGCTTAATCAATCTTCAAACATACATAAATAATTATCATTTCCCACTAGTTT
GGTGCTTTGTGTTTTTAGTTTGAACTTGTCTTTGAACCCTAGCTATAGGATCC
ACCTGATATCTACATGCAAATAAGTGGCACTGCCTTATTAGTTCCATGATGCTC
TATCATTGCACTCACACCACCTTATTTGAGCAGAGGAAATTTAAATGTGTAATT
CATTTTAAAATCACTACTTTGCTAACTAGATTCTCTATATCCAAAAAAAAAA

Fig.2 The cDNA sequence of *Phaseolus coccineus* F3H gene

Note: The boxed parts are in silico cloned sequences longer than RT-PCR results part, which are not included when the primers were designed. The underlined part is RT-PCR results more than electronic cloning sequence. The letters with shades of gray are the start codon of ATG and termination codon of TAA.

1 MAPTAKTLTYLAQEKTLESSFVRDEDERPKVAYNQFQFSDEIPVISLAGIDDVHGRRQSICQKIVQACEDWG
71 IFQVVDHGVDQKLVSEMTELATQFFALPPPEEKLRFDMSGGKKGGFIVSSHLQGESVQDWREIVTYFSYPK
141 KERDYSRWPDTPEGWRAVTEEYSEKLMGLACKLLEVLSEAMGLEKEALTNACVDMDQKVVVNYYPKCPQP
211 DLTLGLKRRHTDPGTITLLLQDQVGGLQATRDNGKTWIIVQPVEGAFVVNLGDHGHFVSNGRFKNADHQAV
281 VNSNHSRLSIATFQNPAPNATVYPLKVRGEKPVLEEPITFAEMYKKKMSKDLEIARMKKLAKEKQLQDL
351 ENEKHLQELDQKAKLETKPLKQILA*

Fig.3 The amino sequence encoded by F3H gene of *Phaseolus coccineus*

3.2 Gene Comparison between F3H cDNA and other species

The alignment of *Phaseolus coccineus*, *Glycine max*, *Medicago truncatula*, *Ipomoea batatas* and *Arabidopsis thaliana* F3H amino acid sequences, constructed using Clustal Omega program, suggested that *Phaseolus coccineus* F3H amino acid sequences was very similar to *Glycine max*. 92% amino acids of F3H in *Phaseolus coccineus* and *Glycine max* were the same (Fig.4).

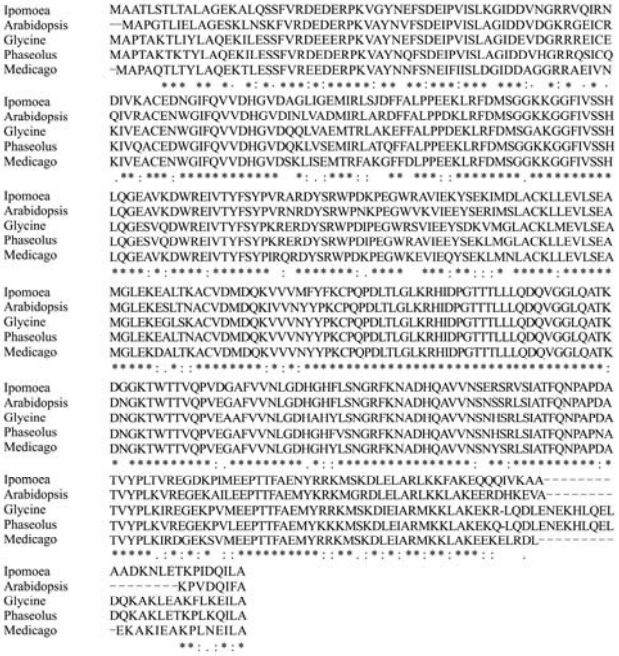

Fig.4 Multi-alignment of *Phaseolus coccineus* F3H amino acid sequences with other plant

Note: The symbols “ * ”, “ : ” and “ . ” represented completely identical, conservative and semi-conservative amino acid residues respectively.

3.3 Secondary structure of *Phaseolus coccineus* F3H protein

The secondary structure of *Phaseolus coccineus* F3H protein was predicted utilizing online service. The results showed that *Phaseolus coccineus* F3H protein had 40.8% α-helix, 14.93% extended strand, 39.73% random coil and 4.53% β- turn.

3.4 Analysis of *Phaseolus coccineus* F3H protein phylogenetic tree

Sequences of *Glycine soja*, *Clitoria ternatea*, *Onobrychis viciifolia*, *Canarium album*, *Dimocarpus longan*, *Litchi chinensis*, *Medicago truncatula*, *Ampelopsis grossedentata*, *Arabidopsis lyrata*, *Brassica napus*, *Ipomoea batatas*, *Malus domestica*, and *Arabidopsis thaliana* were downloaded from protein database. And these sequences were stored in a FASTA file including F3H protein sequence of *Glycine max*, *Phaseolus coccineus*. The Neighbor-Joining phylogenetic tree was constructed with the FASTA file by software

MEGA 5.05（Fig.5）. The protein phylogenetic tree analysis showed *Leguminosae* plants (*Glycine max*, *Glycine soja*, *Clitoria ternatea*, *Phaseolus coccineus*) were clustered into one branch, and *Cruciferae* plants (Arabidopsis thaliana, *Brassica napus*, *Arabidopsis lyrata*) were clustered into one branch. *Phaseolus coccineus* F3H protein was closest to *Glycine max* and *Glycine soja*. The genetic relationship between the F3H proteins was consistent with phylogenetic tree.

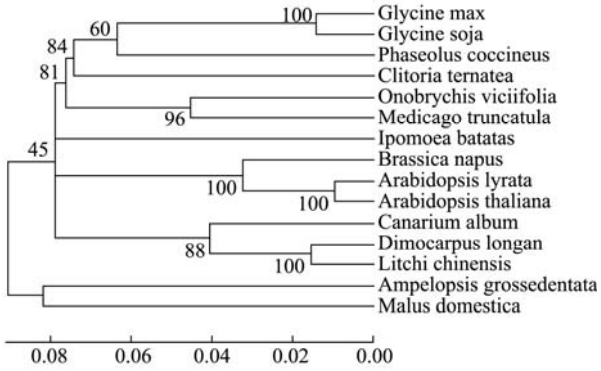

Fig.5 Analysis of *Phaseolus coccineus* F3H protein phylogenetic tree

3.5 Advanced Structure of *Phaseolus coccineus* F3H protein

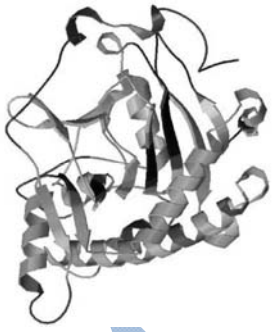

Fig.6 The predicted three-dimensional structure of *Phaseolus coccineus* F3H protein

The structure prediction from primary to advanced structure was an important task in field of the protein research. The three-dimensional structure model of *Phaseolus coccineus* F3H protein was predicted by the Swiss-Model server and homology modeling based on the available structures. The homologous sequence of pdb1gp6A and pdb1w9yA with high homology and three-dimensional structure were two models. Because the C terminal of *Phaseolus coccineus* F3H was poor homology to *Glycine max*, and less than ten amino acids, the template of pdb1w9yA matching 2-342 to *Phaseolus coccineus* F3H was selected for homology modeling (Fig.6). The results were close to the protease real space conformation.

4 Discussion

4.1 In silico cloning is a method developed in recent years for functional gene identification by using genome and EST database [6]. Compared to the traditional methods, such as molecular hybridization, screening of genomic or cDNA library, it is advanced for low cost, high efficiency, easy operation, etc [7]. With more and more EST and genome sequencing data reported, it would become possible and feasible to isolate and identify the functional genes by in silico cloning [8]. Many successful examples strongly support the fact that in silico cloning is absolutely a feasible tool for gene cloning and presents some advantages, compared to the traditional methods [9].

4.2 In this study, we present here in silico cloning and characterization of one flavonoid biosynthetic genes, flavanone-3 hydroxylase of *Phaseolus coccineus*. The F3H transcripts are abundant in *Phaseolus coccineus* seeds. In addition, a sequence similar to the Myb-like box A/CACC T/AAA/CC found in several genes of the flavonoid pathway [10]. And it was thought to be involved in the regulation of the flavonoid biosynthesis pathway, is present in inverse orientation between -1024 and -1016 upstream from the start codon of the *Phaseolus coccineus* F3H gene [11~12]. Future analysis of *Phaseolus coccineus* plants transformed with the F3H promoter-GUS fusions may demonstrate the involvement of the Myb-like sequence in the expression of this gene. The structure and function of *Phaseolus coccineus* F3H were analyzed and predicted using bioinformatics methods successfully. The results revealed that it was a convenient technique for cloning novel gene by searching EST database with homologous gene of model living things. To our knowledge, it was the first report about cloning of *Phaseolus coccineus* F3H cDNA with in silico cloning. This research will provide theory and reference for flavonoid biosynthesis research in *Phaseolus coccineus*.

References

[1] Falginella L, Castellarin S D, Testolin R, et al. Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage [J]. BMC Genomics, 2010, 11: 562-579

[2] Halbwirth H. The creation and physiological relevance of divergent hydroxylation patterns in the flavonoid pathway [J].

Int. J. Mol. Sci., 2010, 11: 595-621

[3] Lepiniec L, Debeaujon I, Routaboul J M, et al. Genetics and biochemistry of seed flavonoids [J]. Annu. Rev. Plant Biol., 2006, 57: 405-430

[4] Han Y, Vimolmangkang S, Soria-Guerra R E, et al. Ectopic expression of apple F3'H genes contributes to anthocyanin accumulation in the Arabidopsis tt7 mutant grown under nitrogen stress [J]. Plant Physiol., 2010, 153: 806-820

[5] Hichri I, Barrieu F, Bogs J, et al. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway [J]. J. Exp. Bot., 2011, 62: 2465-2483

[6] Han W, Ding P, Xu M, et al. Identification of eight genes encoding chemokine-like factor superfamily members 1-8 (CKLFSF1-8) by in silico cloning and experimental validation [J]. Genomics, 2003, 81: 609-617

[7] Nagamatsu A, Masuta C, Matsuura H, et al. Down-regulation of flavonoid 3'-hydroxylase gene expression by virus-induced gene silencing in soybean reveals the presence of a threshold mRNA level associated with pigmentation in pubescence [J]. J. Plant Physiol., 2009, 166: 32-39

[8] Yu F, Wang K, Shao X F, et al. In silico cloning and transcript level analysis of PpPFK in peach fruit stored at different low temperatures [J]. Modern Food Science and Technology, 2013, 29(9): 2125-2130

[9] Toda K, Kuroiwa H, Senthil K, et al. The soybean F3'H protein is localized to the tonoplast in the seed coat hilum [J]. Planta, 2012, 236: 79-89

[10] Zabala G, Vodkin L O. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean W1 locus [J]. Crop Sci., 2007, 47: 113-124

[11] Takahashi R, Dubouzet J G, Matsumura H, et al. A new allele of flower color gene W1 encoding flavonoid 3',5'-hydroxylase is responsible for light purple flowers in wild soybean Glycine soja [J]. BMC Plant Biol., 2010, 10: 155-164

[12] Toda K, Yang D, Yamanaka N, et al. A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color [J]. Plant Mol Biol., 2002, 50: 187-196