

白菜黄烷酮-3-羟化酶基因的电子克隆与序列分析

马光¹, 梁菲菲¹, 郭继平¹, 齐善厚², 仲璐¹

(1. 衡水学院 生命科学系, 河北 衡水 053000; 2. 衡水学院 疾病预防控制中心, 河北 衡水 053000)

摘要: 现根据同源基因相对保守性, 利用已知的油菜 F3H 基因序列, 采用电子克隆的方法获得了白菜 F3H 基因的 cDNA 序列; 并采用生物信息工具对白菜 F3H 的性质包括氨基酸序列组成, 物理和化学性质, 疏水性或亲水性, 二级和三级结构的特点进行了研究。结果表明: 白菜 F3H 基因 cDNA 全长为 1 228 bp, 包含 1 个完整的开放读码框(1 077 bp), 编码 358 个氨基酸。其分子量为 40 104.6 Da, 等电点为 5.45; 其三维结构显示其是具有 7 个螺旋和 12 个折叠和一些不规则卷曲组成的球状结构。

关键词: 黄烷酮-3-羟化酶; 白菜; 电子克隆

中图分类号: S 682.36 **文献标识码:** A **文章编号:** 1001-0009(2012)17-0116-03

黄烷酮-3-羟化酶(F3H; EC 1.14.11.9)是植物花青素生物合成过程中的一个关键酶, 它催化黄烷酮的C-环第3位置羟化形成二氢黄烷醇。其首次被报道是在紫罗兰的粗提物中被发现, 并通过欧芹的细胞培养方式对其性质进行了研究^[1], 随后, F3H 基因从矮牵牛中被克隆并在大肠杆菌中的功能表达中表现出较高活性^[2]。后来, 更多的物种 F3H 基因的序列和特点被报道, 如大麦、海棠、紫花苜蓿、玉米、拟南芥和紫苏^[3]。在不同生物中, F3H 的表达特点有所区别。在矮牵牛和金鱼草中, 类黄酮生物合成途径的前期步骤和后期步骤的调控是不同的。在矮牵牛中 F3H 基因属于前期基因, 而在金鱼草中却属于后期基因。在玉米中, 类黄酮生物合成的整个途径在组织中是通过共调控来产生花色素的, 但是 F3H 基因的表达仅仅同花药中的黄酮醇积累一致。在紫花苜蓿的研究中发现, F3H 在根部和根瘤中也有表达, 但是 F3H 酶在其中发挥的作用还不清楚^[4]。

在芸薹属中, 油菜(*Brassica napus*)的 F3H 基因已经被克隆^[5]。油菜中 F3H 是由 2 个基因编码, 即 F3H-1 基因和 F3H-2 基因。在白菜中, F3H 基因的序列尚未被报道。该研究采用已知的油菜 F3H 基因为探针, 利用已有的中国大白菜(*Brassica rapa* Chinese Cabbage Group 'taxid:51351')的 EST 数据库对白菜 F3H 基因进行电子克隆, 初步获得其 cDNA 编码序列, 并对其序列以及编码蛋白的特点进行研究。

1 材料与方法

1.1 白菜 F3H 基因的电子克隆

分别采用油菜(*Brassica napus*)的 F3H-1 基因(GenBank: DQ288239.1)和 F3H-2 基因(GenBank: DQ513329.1)的 cds 作为探针 Blastn, 搜索白菜(*Brassica rapa* Chinese Cabbage Group 'taxid:51351')的表达序列标签(EST)数据库(<http://www.ncbi.nlm.nih.gov/dbEST/>)。将得到的 EST 序列进行在线拼接(<http://pbil.univ-lyon1.fr/cap3.php>), 将拼接后的较长序列再作为探针重复上一步的 Blastn, 搜索白菜表达序列标签 EST 数据库, 直到拼接序列无法延伸为止。最后得到的拼接序列即为电子克隆得到的白菜 F3H 基因序列。

1.2 白菜 F3H 基因的生物信息学分析

电子克隆得到的白菜 F3H 基因序列提交在线开放读码框寻找软件, 以查找其起始密码子和终止密码子位置。对得到的白菜 F3H 基因编码的蛋白序列采用多序列蛋白质序列在线比对软件 clustalw 2.1(<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)将其和油菜的 F3H-1 以及 F3H-2 基因编码的蛋白进行相似性比对分析。克隆得到的白菜 F3H 基因编码的蛋白的一级结构特点采用在线服务器(<http://www.expasy.org/tools/protparam.html>)进行分析。二级结构分析采用在线服务器(http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)进行。疏水性分析采用 ProtScale 程序(<http://www.expasy.org/cgi-bin/protscale.pl>)进行。三维结构建模采用在线服务器(http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1)进行。

第一作者简介: 马光(1977-), 男, 山东济宁人, 讲师, 研究方向为植物生理学。E-mail: maaohan@163.com.

基金项目: 河北省高等学校自然科学研究计划资助项目(Z2010159)。

收稿日期: 2012-05-17

2 结果与分析

2.1 Blastn 对白菜 EST 数据库的搜索结果

利用 Blastn 程序,以油菜的(*Brassica napus*)的 F3H-1 基因(GenBank: DQ288239.1)和 F3H-2 基因(GenBank: DQ513329.1)的 cds 序列作为种子序列对白菜的表达序列标签数据库进行同源搜索。2 次结果相同,得到有意义的同源性序列 14 条(图 1)。

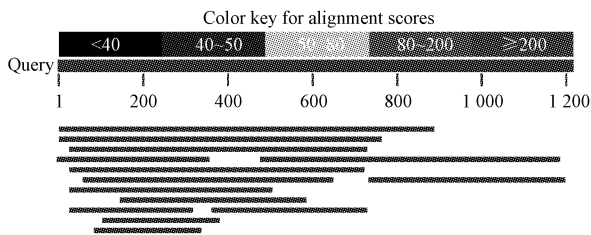


图 1 以油菜的 F3H-1、F3H-2 基因为探针对白菜 EST 数据库进行 Blastn 的搜索结果

2.2 EST 片段的拼接及开放读码框的确定

将 Blastn 得到的 14 条 EST 序列保存在 1 个 FASTA 文件中,提交在线拼接软件 CAP3 进行拼接后得到 1 条 contig(图 2)。将此 contig 作为探针对白菜 EST 数据库进行 Blastn 搜索得到的结果与图 1 中的 14 条序列相同。所以此 contig 即为电子克隆得到的白菜 F3H 基因的 cDNA 编码序列。

```
TTTCATGCTATTATTTTACATCTTGATATAATTATAATGGCTCCAGGAACCTCACTAGCTGAGCTTGGCGGAGAG
ACTAAGCTCAACTCTAAGTTCGTCGCGGAGCAGGATGAACGTCCCAAGGTTCCTCAATGAGTTTA
GCACCGAGATCCCGGTGATCTCCCTCGCCGGAATCGAGATGTGGTGGGAAAGAGGAGAGATCT
GTGACAGATCTGTGAGGCTGTGAGAACTGGGGTGTGTTTCAGGTGCTGATCATGGGTGGATA
CTAGTTTGGTGGCCGATATGACTCTCTGCTCGCTGAGACTTCTGCTTACCTCTCTGAGGAGAACT
CAAGTTCGACATGTCTGGTGGTAAAAAGGAGGATTCATGCTCTAGTCACTCTCAGGAGAGTCT
GTTCAAGACTGGAGAGAGATCTGTCAGCTATTCTCTGACCCGGTGAGAAACAGAGACTACTCAGCG
TGGCCGACTAAGCCGGAAGGATGGTGAAGTACGCGAGGAGTACAGCGAGAGGCTGATGGGTTT
GGCTTGTAAAGCTTCTTGAGGTCTGCTGAGCAACGCGGCTCGAGAAAGAGGCACTACCAATCGC
ATGCGTCGATATGACACAGAAATAGTTGTAATTAATACCAAAATGCCCTCAGCCTGATCTTACCC
TCGCACTCAAGCGTCACACTGACCTGGAACCATCACTTTCGCTCTCAAGACCAAGTCGGTGTGTT
ACAAGCCACAGAGAGATGGGAAGACATGATTACGGTTCAGCCTGTGTAAGGAGCTGTGTGCTGT
TAATCTGGGCAACCATGGTCTACTCTGAGCAACGCGGAGGTTCAGAAACGCTGACCAACAGGCGGT
TGTGAATCCAACTCAGCAGCACTATCAATGACCACTCCAGAAATCCGCGCCGGAAGCAACCGT
GTATCCGCTTAAAGTGAGAGAAGGAGAGAACCGATCTTGGAGGAGCAATTAAGTTTGGGAGAT
GTATAGGAGAAGATGAGTAGAGATATCGAGCTGCTGCTCAAGAAGCTGGCGAAGAAAGAAC
TGACCACAAGGAAGCTGCCAAGCTCTAGACCAATATCGCTTAGACTCTCTGTGTGGCCCTTACT
ATTGTGTGGTTCGTTGTGTAAGTCTGACTGAAATAGGGTATGTCATGTGTGGTGAATTTAAGTTTAT
ATCGTTTCTATAATATAAAA
```

图 2 电子克隆的白菜 F3H 基因 cDNA 序列

利用 NCBI 网站的开放读码框搜寻软件 ORF finder 对在线拼接成的 F3H 基因的 cDNA 序列进行分析。得到 1 个长度为 1 077 bp 的开放读码框(图 2 中下划线的部分),编码 358 个氨基酸。

2.3 CLUSTALW 结果

将油菜的 F3H-1(BnF3H-1)、F3H-2(BnF3H-2)和拼接得到的白菜 F3H 编码蛋白(contig)利用 Clustalw 2.1 进行同源比对。结果表明,白菜的 F3H 基因(图 3 中的 contig)与油菜 F3H-1 基因编码的蛋白同源性更高,仅有

1 处不同(254 位),同源性为 99.72%。与油菜的 F3H-2 基因同源性较低,有 4 处不同(7、51、324、331 位),同源性为 98.88%。

```
contig  MAPGTLTELAGETKLNKSFVRDEDERPKVAYNEFSTEIPVISLAGIDDVGGKRGECRQI 60
BnF3H-1 MAPGTLTELAGETKLNKSFVRDEDERPKVAYNEFSTEIPVISLAGIDDVGGKRGECRQI 60
BnF3H-2 MAPGTLTELAGETKLNKSFVRDEDERPKVAYNEFSTEIPVISLAGIDDVGGKRGECRQI 60
*****

contig  VEACENWGVFQVVDHGVDTSLVADMTLRDLFFALPPEEKLKFDMSGGKGGFVSSHQ 120
BnF3H-1 VEACENWGVFQVVDHGVDTSLVADMTLRDLFFALPPEEKLKFDMSGGKGGFVSSHQ 120
BnF3H-2 VEACENWGVFQVVDHGVDTSLVADMTLRDLFFALPPEEKLKFDMSGGKGGFVSSHQ 120
*****

contig  GESVQDWREIVTYSYPVRNDRYSRWPTKPEGVVKVTEEYSERLMGLACKLLEVLSEAMG 180
BnF3H-1 GESVQDWREIVTYSYPVRNDRYSRWPTKPEGVVKVTEEYSERLMGLACKLLEVLSEAMG 180
BnF3H-2 GESVQDWREIVTYSYPVRNDRYSRWPTKPEGVVKVTEEYSERLMGLACKLLEVLSEAMG 180
*****

contig  LEKEALTACVDMQKIVVNYPKCPQDLTLGLKRHTDPGTITLLQDQVGGIQAATRDD 240
BnF3H-1 LEKEALTACVDMQKIVVNYPKCPQDLTLGLKRHTDPGTITLLQDQVGGIQAATRDD 240
BnF3H-2 LEKEALTACVDMQKIVVNYPKCPQDLTLGLKRHTDPGTITLLQDQVGGIQAATRDD 240
*****

contig  GKTWITVQPEGAVVVNLGDHGHYLSNGRFKNADHQAIVNSNSSRLSIATFQNPAPAEATV 300
BnF3H-1 GKTWITVQPEGAVVVNLGDHGHYLSNGRFKNADHQAIVNSNSSRLSIATFQNPAPAEATV 300
BnF3H-2 GKTWITVQPEGAVVVNLGDHGHYLSNGRFKNADHQAIVNSNSSRLSIATFQNPAPAEATV 300
*****

contig  YPLKVRGEKPILEPITFAEMYRRKMSRDIELARLKLKEEHDHKEAAKPLDQIIA 358
BnF3H-1 YPLKVRGEKPILEPITFAEMYRRKMSRDIELARLKLKEEHDHKEAAKPLDQIIA 358
BnF3H-2 YPLKVRGEKPILEPITFAEMYRRKMSRDIELARLKLKEEHDHKEAAKPLDQIIA 358
*****
```

图 3 白菜 F3H 基因与油菜 F3H-1、F3H-2 蛋白的同源性比较

2.4 白菜 F3H 的一级结构特点

白菜 F3H 蛋白的一级结构特点分析结果表明,其由 20 种氨基酸组成。亮氨酸含量最高(8.9%),色氨酸和半胱氨酸最低(1.4%),分子量为 40 104.6 Da,等电点为 5.45,分子式为 $C_{1776}H_{2813}N_{489}O_{542}S_{13}$,总共有 5 633 个原子组成。消光系数为 42 650(280 nm)。估计的半衰期是 30 h(哺乳动物网状细胞,体外)。热稳定指数(II)为 40.71,表明其为不稳定蛋白。脂溶指数(Aliphatic index)74.10。总平均亲水性为-0.438。

2.5 白菜 F3H 的二级结构特点

白菜 F3H 二级结构预测结果表明,其含有 37.71%的 α -螺旋,15.43%延伸链,42.29%无规卷曲和 4.57%的 β 折叠。对其疏水性分析结果表明,344 位 Asp 和 345 位 His 亲水性最高(-2.933)和 42 位 Ser 疏水性最高(2.278)(图 3)。综合各氨基酸的结果可以得出结论白菜 F3H 为一个亲水性蛋白质。

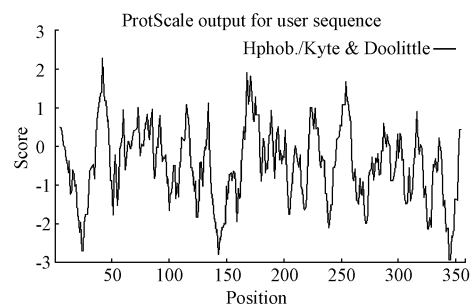


图 4 白菜 F3H 蛋白亲水性分布

2.6 白菜 F3H 的三级结构建模

以蛋白 Parent PDB [1gp6A](1.75 Å) 为模板对白菜 F3H 蛋白进行了同源建模,结果见图 5。由图 5 可知,白菜 F3H 蛋白近似球形。由 7 个螺旋和 12 个折叠和一些无规则卷曲组成,形成了紧凑的球形空间结构以发挥其生理活性。



图 5 白菜 F3H 的三维结构

3 讨论与结论

电子克隆是最近开发的一种利用基因组和 EST 数据库来研究基因功能的新方法,与传统实验室克隆方法(如分子杂交,构建基因组文库,EST 文库筛选)相比,电子克隆具有低成本、高效率、易操作等优势^[6]。随着 EST 数据和基因组测序结果数量的增加,它的应用将越来越广泛。在该研究中,通过 EST 数据库搜索和拼接获得了白菜 F3H 基因的序列,采用 F3H-1 和 F3H-2 的搜

索和拼接结果完全一致表明了 2 种可能,1 种是白菜中编码 F3H 蛋白的基因只是 1 种;另 1 种可能是白菜的 EST 数据库的来源对白菜不同器官或者不同发育时期划分的还不够详细。对拼接序列编码蛋白的同源比较和性质研究的结果表明,采用电子克隆方法初步获得了白菜 F3H 基因的序列。由于白菜 F3H 基因尚未见报道,该研究获得的序列对通过分子生物学手段获得白菜 F3H 基因并研究其功能提供了有力支持。

参考文献

- [1] Forkmann G, Heller W, Grisebach H. Anthocyanin biosynthesis in flowers of *Matthiola incana*. Flavanone 3- and flavonoid 3'-hydroxylases[J]. Z Naturforsch Sect C Biosci, 1980, 35: 691-695.
- [2] Britsch L, Ruhnau-Brich B, Forkmann G. Molecular cloning, sequence analysis, and *in vitro* expression of flavanone 3 beta-hydroxylase from *Petunia hybrida*[J]. J Biol Chem, 1992, 267: 5380-5387.
- [3] Shen G A, Pang Y Z, Wu W S, et al. Cloning and Characterization of a Flavanone 3-Hydroxylase Gene from *Ginkgo biloba*[J]. Biosci Rep, 2006, 26: 19-29.
- [4] Xu B B, Li J N, Zhang X K, et al. Cloning and molecular characterization of a functional flavonoid 3'-hydroxylase gene from *Brassica napus*[J]. J Plant Physiol, 2007, 164: 350-363.
- [5] 马光. 芜菁花青素合成关键基因 F3H 启动子区序列及蓝光受体 CRY1 基因的克隆与初步功能分析[D]. 哈尔滨: 东北林业大学, 2008.
- [6] Zhang H M, Jiang M G, Feng Y J. In silico cloning of MgEno-1 cDNA from *Magnaporthe grisea*[J]. China J Bioinformatics, 2005(4): 57-61.

Silico Cloning and Sequence Analysis of F3H Gene of Chinese Cabbage

MA Guang¹, LIANG Fei-fei¹, GUO Ji-ping¹, QI Shan-hou², ZHONG Lu¹

(1. Department of Life Sciences, Hengshui College, Hengshui, Hebei 053000; 2. Center for Disease Control and Prevention, Hengshui College, Hengshui, Hebei 053000)

Abstract: Flavanone 3-hydroxylase (F3H), a critical enzyme, catalyzes an step in anthocyanin biosynthesis. According to the conservation of homologous gene, with F3H gene in *Brassica napus*, the cDNA of F3H gene in Chinese cabbage was in silico cloned. Some characters of the F3H, such as the composition of amino acid sequences, physical and chemical properties, hydrophobicity or hydrophilicity, secondary and tertiary structure of the protein and function were analyzed and predicted by the tools of bioinformatics. The results showed that the length of cloned F3H cDNA was 1 228 bp and it contained a complete ORF (1 077 bp), encoded 358 amino acids. The calculated molecular weight of it was 40 104.6 Da, theoretical pI of 5.45. It was a bulbous structure with 7 helices, 12 sheets, and some irregular coils in tertiary structure.

Key words: Flavanone 3-hydroxylase; Chinese cabbage; in silico cloning