

# 植物查尔酮异构酶的生物信息学分析

雷 桅<sup>1</sup>, 邹 祥<sup>2</sup>, 向 阳<sup>1</sup>, 汤绍虎<sup>1</sup>, 孙 敏<sup>1</sup>

(1. 西南大学 生命科学学院 三峡库区生态环境教育部重点实验室, 重庆 400715; 2. 西南大学药学院, 重庆 400715)

**摘 要:** 采用生物信息学方法和工具对 GenBank 中的洋葱、豌豆、番茄和茶等植物的黄酮类化合物合成关键酶查尔酮异构酶(CHI)的核酸和氨基酸序列进行了对比、分析和建模, 进而对其分子结构、理化性质、亚细胞定位、蛋白转运肽、跨膜结构域、疏水性、分子系统进化、蛋白质二级和三级结构等重要参数进行了预测和推理。结果表明: 该类酶基因的全长包括 5'、3'非翻译区和一个开放阅读框, 无蛋白转运肽, 且定位于细胞质基质, 是一个疏水性蛋白, 二级结构均以随机卷曲和  $\alpha$ -螺旋为主要构件, 洋葱和豌豆的 CHI 三维建模成功。

**关键词:** 查尔酮异构酶; 生物信息学; 黄酮类化合物

**中图分类号:** Q 946.5; Q 558 **文献标识码:** A **文章编号:** 1001-0009(2008)02-0193-05

黄酮类化合物是由苯丙烷类化合物衍生得到的一大类植物次生代谢产物, 其特点是具有 C6-C3-C6 的基本骨架, 并可根据中间吡喃环的不同氧化水平和两侧 A、B 环上连接的各种取代基, 而分为不同的黄酮类型。研究证明黄酮类化合物对植物自身进行正常的生理活动起着重要作用, 如调节生长、UV 防护、抗虫抗病、花色形成, 影响花粉育性, 诱导植物根部与共生菌相互作用等。因此作为园艺植物中一类重要的化学成分, 一直备受农艺界的广泛关注, 是花色基因工程和天然产物代谢工程研究的一类主要目标产物。

查尔酮异构酶(chalcone isomerase, CHI)是第 1 个被认识的黄酮类化合物合成相关酶, 也是黄酮代谢途径中的关键酶之一, 它催化柚皮苷查尔酮异构化形成有生物活性的二羟基黄酮。CHI 一般可分为两种类型, 一种主要存在于豆科植物中, 催化 6'-脱氧查尔酮和 6'-羟基查尔酮生成异黄酮类化合物和黄酮类化合物; 另一种主要存在于非豆科植物中, 催化 6'-脱氧查尔酮生成 5'-羟基黄酮<sup>[1]</sup>。目前 CHI 基因已先后从豌豆、矮牵牛、苜蓿等多种植物和人类粪便厌氧细菌中被克隆出来<sup>[2-5]</sup>, 通过代谢工程的研究证明改变 CHI 活力可以有效调控黄酮类化合物的含量, 例如超量表达牵牛花 CHI 基因使番茄果皮中的黄酮水平提高了 78 倍<sup>[6]</sup>, 降低康乃馨中的 CHI 酶活得到奇特的黄色花朵<sup>[7]</sup>, 而 CHI 失活

的洋葱突变体中槲皮素水平明显下降<sup>[8]</sup>。因此对 CHI 的研究是当前一个十分重要的热门话题。现利用生物信息学的方法, 以洋葱(*Allium cepa*)为重点, 对豌豆(*Pisum sativum*)、番茄(*Lycopersicon esculentum*)、茶(*Camellia sinensis*)等园艺植物查尔酮异构酶的基因及相应氨基酸序列的理化性质、结构特征、生化功能及系统演化关系等进行预测和分析, 以期今后深入研究 CHI 家族蛋白的酶学特性和黄酮生物合成的分子机理提供理论依据。

## 1 材料与方法

数据来源于 NCBI 核酸和蛋白质数据库中已登录的查尔酮异构酶(CHI)的核酸序列及其对应的氨基酸序列: 洋葱(Accession: AY541034)、豌豆(Accession: U03433)、番茄(Accession: AY348871)、茶(Accession: DQ120521)。

利用 Vector NTI 8、Clustal X、MEGA3、ViewerLite 4.2 软件和 www.ncbi.nlm.nih.gov、www.expasy.org 等网站提供的各种在线生物信息学工具进行预测和分析。核酸及氨基酸序列的分子结构和理化性质分析、开放阅读框的查找和翻译使用 Vector NTI 8 软件完成; 核酸及氨基酸序列的同源性比对使用 Clustal X 软件和 NCBI 上的 Blast 在线工具完成; 分子系统发生树使用 MEGA3 的 Phylogeny 中的 Neighbor-Joining 方法构建; 亚细胞定位情况的考察使用 TargetP 1.1 Server 在线完成, 并对其结果进行 SignalP 及 ChloroP 分析, 得到蛋白质信号肽和叶绿体转运肽的相关信息; 跨膜结构域和疏水性分析分别使用 TMHMM 和 ProtScale 完成; 蛋白质二级结构预测和三级建模分别使用 GOR 和 SWISS-MODEL 完成。

## 2 结果与分析

**第一作者简介:** 雷桅(1982-), 男, 硕士研究生, 主要从事药用植物代谢工程研究。E-mail: weil06@swu.edu.cn

**通讯作者:** 孙敏。

**基金项目:** 三峡库区生态环境教育部重点实验室开放基金资助(EF200609)。

**收稿日期:** 2007-09-01

## 2.1 植物 CHI 基因的核酸及相应氨基酸序列的分子结构和理化性质分析

Vector NTI 8 分析得到的洋葱、豌豆、番茄和茶的核酸序列和氨基酸序列相关结构和性质数据见表 1, 其中洋葱 CHI 基因的全长 mRNA 序列及其推导的氨基酸序列如图 1 所示。由表 1 可见, 不同植物 CHI 基因全长、开放阅读框碱基数及其所编码的氨基酸残基数相差较小, 且编码区的起始密码子均为 ATG, 而终止密码子则不一致, 洋葱和豌豆的是 TGA, 番茄和茶的是 TAA。同时, 不同植物 CHI 的分子量、等电点、摩尔消光系数、酸性氨基酸比例、碱性氨基酸比例、带电氨基酸比例、极性氨基酸比例、疏水性氨基酸比例等理化指标均表现出一定的差异, 这表明 CHI 基因可能存在一定的物种特异性和生化多态性; 含量最丰富的氨基酸也不完全相同, 但至少都含有 Lys, 说明 Lys 在 CHI 的催化反应中起着重要作用。

表 1 植物 CHI 基因的核酸及相应氨基酸序列的分子结构和理化性质分析

指标	洋葱	豌豆	番茄	茶
基因全长/bp	905	902	1 001	1 002
开放阅读框全长/bp	678	669	786	720
起始位点及密码子/bp	85 ATG	22 ATG	81 ATG	92 ATG
终止位点及密码子/bp	763 TGA	691 TGA	867 TAA	812 TAA
氨基酸残基数	226	223	262	240
分子量/kDa	24.170	24.787	28.972	26.353
等电点	5.10	7.20	5.65	5.34
摩尔消光系数	16 740	11 410	7 800	7 920
	Arg 8.85	Lys 8.07	Lys 13.36	Lys 12.08
含量最丰富的氨基酸/%	Glu 8.41	Ile 8.07	Glu 10.69	Leu 9.17
	Gly 9.73	Leu 11.66	Gly 7.63	Gly 8.33
酸性氨基酸比例/%	12.83	13.45	16.03	16.25
碱性氨基酸比例/%	9.73	13.45	15.27	14.58
带电氨基酸比例/%	26.55	34.08	33.97	34.58
极性氨基酸比例/%	22.12	21.52	20.61	20.00
疏水性氨基酸比例/%	38.05	35.43	31.30	31.25

```

1  egcggg aagcag tegt tca acccag agta cgcggg gact acat cat tt cta gatt tt cat cta tacc teat aat l ca agaa
    M E A V T K L D V E G T A F D S V I I P P G S S K T H F L G G A G
84  [ATG] GAAGCAGTGC AAAAGTTGGACGTAG AAGGAACTGCCTTTGATTCA GTCATCATCCCTCCCGGTTATCCAAAACGCAC TTCTCGGCGGTCAGGT
    V R G L E I G G K F I A F T A I G I Y L E T D S I P F L A D K W K
184  GTAAAGGGT TTAGAA AATTGAGGTAAAT TTATAGCGT TTAGCGCAATCGGCA TCAC TTGAAACAGATTC AAT TCCGTTTC TTGCTGATAAATG GAAA
    G K T G E E L A G S L D F F R D I C T G P F E K F T N V T M I L P
242  GGAAAACAGCGGAGGAACTCGCTCGT TCACTGCACTTT TCGAGATATATG CACAGGACCC TTGAGAAATTTACTA ATG TAA CAATGATTCTCCCT
    L T G E Q Y S E K V T E N C V A Y W K A I G I Y T D A E A S A V D
359  CTAAGGGAGAACACTACTCCGAAAAGTAACAGAAAAT TGTGTGCTTAT TGGAAAGCAAT TGGAAATCTACACAGATGCAGAACCGTCCGCTGTTGAT
    K F K Q A F K P E S F P P G S S I L F T H T P S G T L K I A F S K
407  AAGTTTAAACAAGCTTTTAAACCTGAGAGTTTCCCTCCTCGTTCATCCATTC TCTTCAC TACATCCCTAGCCGGAACACTAAGATTGCATTTCAAA
    D G S V S K D E G V L I E N K A L T Q A V L E S I I G E H G V S P
580  GATGGTTCGGT TTTCTAAAGATGAAGCGGTACTCATTGAGAATAAGCGGCTGACACAAGCCGATTGGAGTCCGAT TATTGGGGAGCATGGTG TATCACCT
    A A K L S I A S R L S E I M N K V G N V E E K L P V L S *
679  GCTGCTAAGCTAAGCATA GCATCGAGATTGTCAGAGATTATGAACAAAGTTGGAAACGTGGAAGAAAAGTTACCGGTGCTTTC tgaattgggaggatt
779  cgtgaaggaa atgatggtctgtgttttaaa ct agatga agtgcactctctgtgtgagatagtc actggagaa tta ctgggttgca ataa ta catlta at
879  tctgtga tcaacaa aaaaa aaaaa
    
```

图 1 洋葱 CHI 的 cDNA 全长序列及推测的氨基酸序列

注: 编码区和翻译的氨基酸用大写字母表示, 起始密码子用 □ 表示, 终止密码子用 \* 表示, UTR 用小写字母表示。

## 2.2 植物 CHI 基因的核酸序列及相应氨基酸序列的比对分析

用 NCBI 网站中的 Blast 程序<sup>[9]</sup>对洋葱 CHI 的核酸和氨基酸序列进行同源性比对, 结果表明洋葱 CHI 的核酸序列与部分植物的 CHI 核酸序列有较高的一致性, 包括茄子(一致性: 70%)、拟南芥(一致性: 70%)、葡萄(一致性: 71%)、烟草(一致性: 72%)、香石竹(一致性: 74%)和百脉根(一致性: 75%)等, 而进一步做蛋白质-蛋白质的 Blastp 比对, 洋葱 CHI 的氨基酸序列与上述植物 CHI 的一致性分别达到 56%、38%、61%、12%、7% 和 55%。对豌豆、番茄和茶进行此类的同源性分析, 也得到相似的规律。由此可见, 植物 CHI 的氨基酸序列比对

比核酸序列比对有更高的相似度。

选取洋葱等 8 条植物 CHI 氨基酸序列, 在 Clustal X 软件<sup>[10]</sup>中进行多重比对分析。分析结果用不同颜色显示(图 2), 颜色越深的表示同源性越高, 最深的区域则表示可能存在的重要功能域。分析结果显示植物 CHI 具有很高的同源性, 在近 C-端催化区域的序列趋于保守, 而近 N-端的序列则差异较大, 这一非催化域往往与蛋白质的亚细胞定位等物种特异性功能相关。

## 2.3 植物 CHI 基因的亚细胞定位分析和蛋白质转运肽的预测和分析

TargetP 1.1 Server 在线工具分析<sup>[11]</sup>结果显示, 洋葱、豌豆、番茄和茶的 CHI 均定位于细胞质基质, 且均无

转运肽, 预测可靠性等级分别为 2、5、2 和 3。可见植物 CHI 在游离核糖体上起始合成后, 不进行蛋白转运, 而

是继续保留在细胞质基质中, 直接催化柚皮苷查尔酮异构化形成二羟基黄酮。

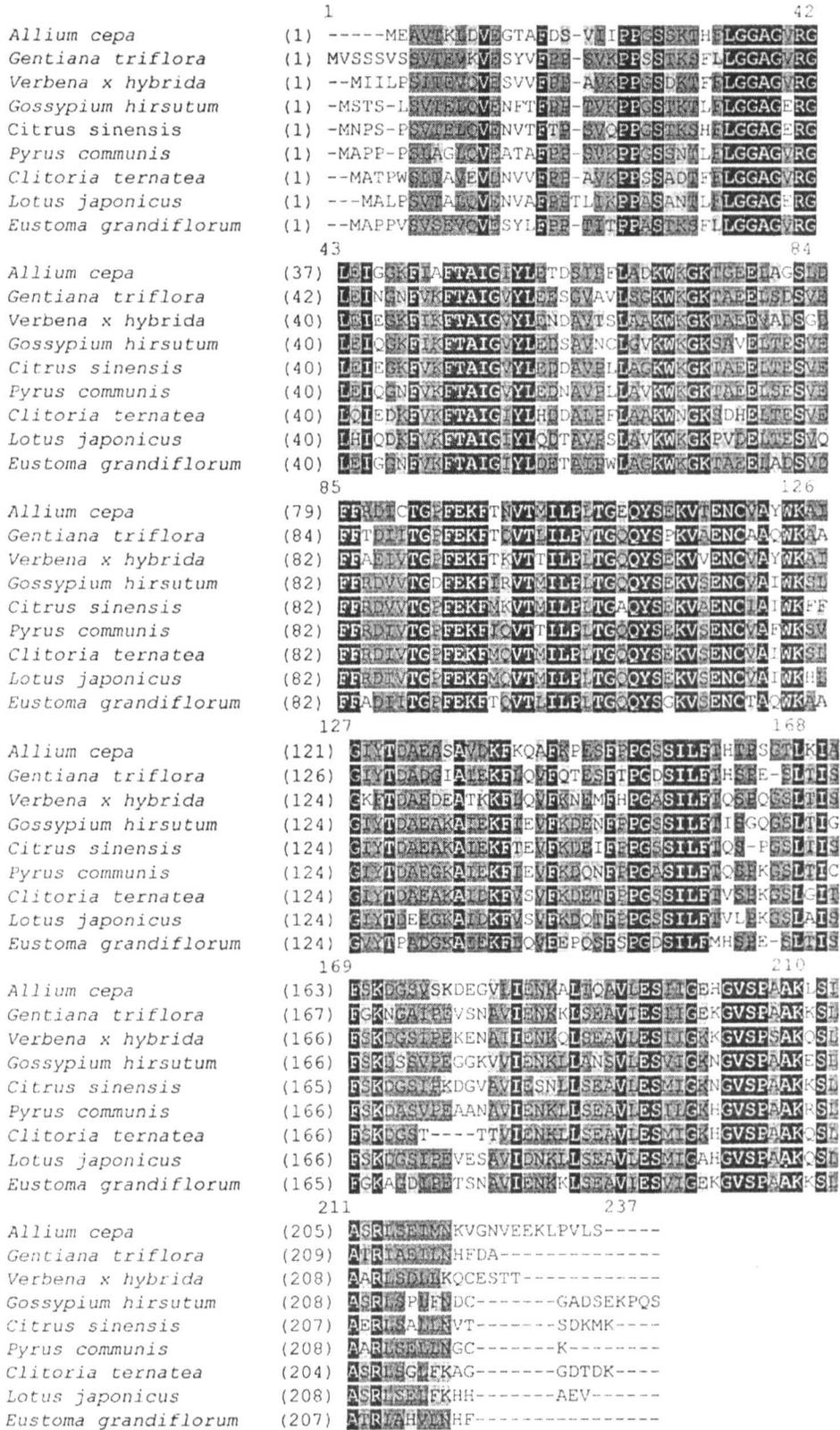


图 2 来自不同植物 CHI 氨基酸序列多重比对

2.4 植物 CHI 基因的跨膜结构域和疏水性预测和分析 用 TMHMM Server v. 2.0 对洋葱 CHI 氨基酸序

列的跨膜结构域进行预测<sup>[2]</sup>, 结果如图 3 所示, 洋葱 CHI 整条肽链都位于细胞膜外, 说明洋葱 CHI 不存在跨

膜区。对豌豆、番茄和茶的 CHI 氨基酸序列的跨膜结构域进行预测分析也得到与此一致的结果, 即 CHI 不存在跨膜结构域, 结合上述转运肽的预测, 可以推断, 植物 CHI 在细胞质基质中合成后, 不经蛋白转运, 直接锚定于细胞质基质中的特定部位行使催化功能。

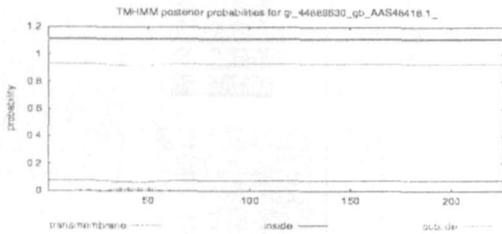


图3 洋葱 CHI 跨膜结构域的预测

由 ProtScale 进行洋葱氨基酸序列的疏水性预测<sup>[3]</sup>, 分析结果(图4)可以看出, 多肽链第20位的 Arg 具有最低分值-4.500, 亲水性最强; 而第200位的 Val 具有最高分值4.200, 疏水性最强。整条多肽链表现为疏水性, 但没有明显的疏水区域。对豌豆、番茄和茶的 CHI 氨基酸序列的疏水性进行预测分析也得到与此一致的结果, 即 CHI 整条多肽链表现为疏水性, 但没有明显的疏水区, 结合跨膜结构域的预测结果, 可以推断, 植物 CHI 不存在明显的疏水区域, 与 CHI 不存在跨膜结构域的特征相吻合。

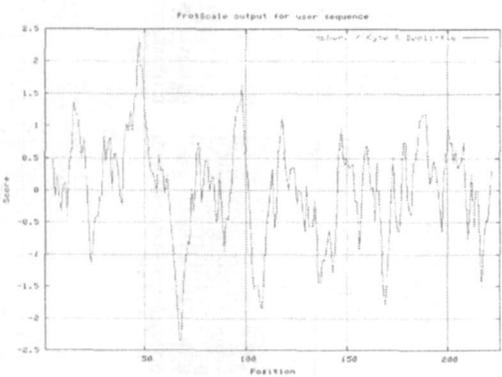


图4 洋葱 CHI 疏水性的预测

### 2.5 植物 CHI 氨基酸序列的分子系统进化树分析

将19条有代表性的 CHI 氨基酸序列在 Clustal X 软件中进行完全比对后, 在 MEGA3 软件中按照 Neighbor-Joining 方案进行分子进化树分析<sup>[4]</sup>, 获得 CHI 氨基酸序列的分子进化树分析结果(如图5)。19种植物的 CHI 氨基酸序列聚成7支: 棉花, 橙聚成第1支; 梨, 牛奶子聚成第2支; 蝶豆, 百脉根聚成第3支; 拟南芥, 萝卜聚成第4支; 洋葱, 红叶美人蕉聚成第5支; 烟草, 美女樱,

菊花, 桔梗, 三花龙胆聚成第6支; 中国大豆, 豌豆, 紫花苜蓿聚成第7支。该进化树反应了这19种不同植物之间的亲缘关系, 如拟南芥和萝卜同属十字花科植物, 聚为一支, 洋葱(百合科, *Liliaceae*)与红叶美人蕉(美人蕉科, *Cannaceae*)虽不同科, 但亲缘关系较近, 也聚为一支。由此可见, 依据氨基酸序列构建的分子进化树虽不能完全真实地反映不同植物之间的系统分类学关系和自然演化进程, 但其结果对于明确判断植物间的亲缘关系仍能提供一定的参考。

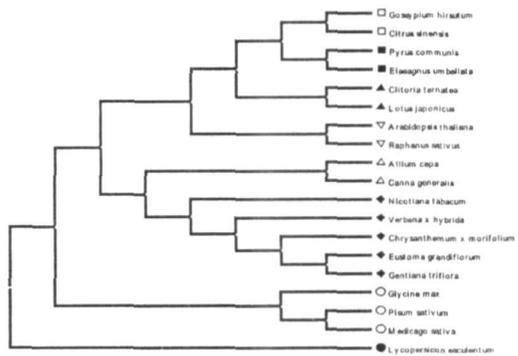


图5 基于 MEGA3.0 中 NJ 方法的洋葱和其它物种的分子进化树聚类分析

### 2.6 植物 CHI 氨基酸序列的二级和三级结构预测和分析

GOR4 在线工具预测洋葱 CHI 氨基酸序列的二级结构<sup>[5]</sup>, 结果如图6所示, 洋葱 CHI 多肽链中, 随机卷曲占 48.67%,  $\alpha$ -螺旋占 28.32%, 延伸链占 23.01%。对豌豆、番茄和茶的 CHI 氨基酸序列进行二级结构预测分析, 得到与洋葱一致的结果, 即随机卷曲和  $\alpha$ -螺旋是 CHI 多肽链中的主要结构元件, 延伸链散布于整个蛋白质中。



图6 洋葱 CHI 二级结构预测结果

注: ▨表示  $\alpha$  螺旋, ▩表示延伸链, □表示随机卷曲

将洋葱 CHI 氨基酸序列上传到 SWISS-MODEL 的建模服务器中进行 CHI 结构的三维建模<sup>[16-17]</sup>, 然后在 ViewerLite 4.2 软件中进行序列编辑, 获得 CHI 的三级结构模型(图7)。据此对豌豆、番茄和茶的 CHI 氨基酸序列进行类似的三维建模, 只有豌豆建模成功(图8), 因为研究使用的 SWISS-MODEL 方案是基于同源建模的蛋白质结构服务器, 即待测氨基酸序列在蛋白质结构数据库中寻找其同源模板, 再将同源蛋白质的结构优化构建出预测的蛋白质三维结构模型。由于服务器中不存在番茄和茶 CHI 氨基酸序列的同源模板, 故建模失败。



图7 洋葱CHI编码蛋白的三维结构 图8 豌豆CHI编码蛋白的三维结构

### 3 小结与讨论

应用生物信息学方法对已知植物黄酮类化合物合成关键酶查尔酮异构酶的核酸和蛋白质序列进行比对、分析、建模,从而对其结构和功能进行推断和预测,对于进一步利用相应的试验手段研究植物查尔酮异构酶的理化性质和分子生物学性质,对物种间CHI基因家族同源性和遗传相似性的分析,有助于新CHI基因和相关EST的发现;而蛋白定位和功能结构域的推断,为针对特定性状遗传改良的基因工程和代谢工程操作,并最终获得观赏性和实用性俱佳的园艺植物新品种都有积极贡献。并且这种基于生物学试验数据与计算技术相结合交叉学科已逐渐成为后基因组时代的重要方法和前沿领域,也是当今生物学发展的必然潮流。

#### 参考文献

- [1] Shimada N, Aoki T, Sato S et al. A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus* [J]. *Plant Physiology*, 2003, 131: 941-951.
- [2] Wood A J, Davies E. A cDNA encoding chalcone isomerase from aged pea epicotyls [J]. *Plant Physiology*, 1994, 104(4): 1465-1466.
- [3] Chen P Y, Wang C K, Soong S C et al. Complete sequence of the binary vector pB1121 and its application in cloning T-DNA insertion from transgenic plants [J]. *Mol. Breed*, 2003, 11: 287-293.
- [4] McKhann H L, Hirsch A M. Isolation of chalcone synthase and chalcone isomerase cDNAs from alfalfa (*Medicago sativa* L.): highest transcript levels

- occur in young roots and root tips [J]. *Plant Mol. Biol*, 1994, 24(5): 767-777.
- [5] Herles C, Braune A, Blaut M. First bacterial chalcone isomerase isolated from *Eubacterium ramulus* [J]. *Arch Microbiol*, 2004, 181: 428-434.
- [6] Shelagh R Muis, Geoff J Collins, Susan Robinson, et al. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols [J]. *nature biotechnology*, 2001, 19: 470-474.
- [7] Kim S, Jones R, Yoo K S, et al. Gold color in onions (*Allium cepa*): a natural mutation of the chalcone isomerase gene resulting in a premature stop codon [J]. *Mol Genet Genomics*, 2004, 272(4): 411-419.
- [8] Takamura T, Tomihama T, Miyajima I. Inheritance of yellow-flowered characteristic and yellow pigments in diploid cyclamen persicum cultivars [J]. *Scientia Hort*, 1995, 64: 55-63.
- [9] Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein data-base search programs [J]. *Nucleic Acids Research*, 1997, 25(17): 3389-3402.
- [10] Thompson J D, Gibson T J, Plewniak F, et al. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. *Nucleic Acids Research*, 1997, 25: 4876-4882.
- [11] Kristin Ehrbar, Siegfried Hapfelmeier. InvB is required for type III dependent secretion of sopA in *Salmonella enterica* serovar typhimurium [J]. *Journal of Bacteriology*, 2004, 186(4): 1215-1219.
- [12] Ikeda M, Arai M, Lao D M. Transmembrane Topology Prediction Methods: A Reassessment and Improvement by a consensus method using a dataset of experimentally characterized transmembrane topologies [J]. *In Silico Biol*, 2002, 2(1): 19-33.
- [13] Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein [J]. *J Mol Biol*, 1982, 157(6): 105-132.
- [14] Saito N, Nei M. The neighbor joining method: a new method for reconstructing phylogenetic trees [J]. *Molecular Biology and Evolution*, 1987(4): 406-425.
- [15] Combet C, Blanchet C, Geourjon C, et al. NPS @ Network Protein Sequence Analysis [J]. *Trends in Biochemical Sciences*, 2000, 25(3): 147-150.
- [16] Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: an automated protein homology-modeling server [J]. *Nucleic Acids Research*, 2003, 31(13): 3381-3385.
- [17] Guex N, and Peitsch M C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling [J]. *Electrophoresis*, 1997, 18: 2714-2723.

## A Bioinformatic Analysis on Chalcone Isomerase in Plant

LEI Wei<sup>1</sup>, ZOU Xiang<sup>2</sup>, XIANG Yang<sup>1</sup>, TANG Shao-hu<sup>1</sup>, SUN Min<sup>1</sup>

(1. Key Laboratory of Eco-environments in Three Gorges Reservoir Region, Ministry of Education, College of Life Science, Southwest University, Chongqing 400715; 2. College of Pharmaceutical Science, Southwest University, Chongqing 400715, China)

**Abstract:** The nucleic acid sequences and amino acid sequences of chalcone isomerase from *Allium cepa*, *Pisum sativum*, *Lycopersicon esculentum* and *Camellia sinensis*, which were registered in GenBank, were analyzed and predicted by the tools of bioinformatics in the following aspects: the composition of nucleic acid sequences and amino acid sequences, molecular structure, physical and chemical characters, subcellular localization, transitpeptides, transmembrane topological structure, hydrophobicity, molecular phylogenetic evolution, secondary and tertiary structure of protein and so on. The results as following: the full-length gene of CHI contains an opening reading frame, 5'-untranslated region and 3'-untranslated region, all lie in cytoplasm, and haven't transmembrane topological structure, were a hydrophobic protein.  $\alpha$ -helix and random coil were main component all secondary structures. The 3D models of *Allium cepa* and *Pisum sativum* are constructed.

**Key words:** Chalcone isomerase; Bioinformatics; Flavonoids