

植物萜类合成酶 3-羟基-3-甲基戊二酰 辅酶 A 还原酶的生物信息学分析

李 嵘, 王喆之

(陕西师范大学 生命科学学院, 陕西 西安 710062)

摘 要: 采用生物信息学的方法和工具对已在 GenBank 上注册的橡胶、烟草、辣椒、穿心莲等植物的萜类合成酶 3-羟基-3-甲基戊二酰辅酶 A 还原酶的核酸及氨基酸序列进行分析, 并对其组成成分、信号肽、跨膜拓扑结构域、疏水性/亲水性、蛋白质二级及三级结构、分子系统进化关系等进行预测和推断。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, 无信号肽, 是一个跨膜的亲水性蛋白, 包括两个功能 HMG-CoA 结合 motif 及两个功能 NADPH 结合 motif, α -螺旋和不规则盘绕是蛋白质二级结构最大量的结构元件, β -转角和延伸链散布于整个蛋白质中, 蛋白质的功能域在空间布局上折叠成“V”形, “V”形的两臂由螺旋状的 N 结构域和 S 结构域构成, 中间部分由 L 结构域构成。

关键词: 植物萜类合成; 3-羟基-3-甲基戊二酰辅酶 A 还原酶(HMGR); 生物信息学

中图分类号: Q942.6, Q943.2 **文献标识码:** A **文章编号:** 1000-3142(2006)05-0464-10

A bioinformatics analysis on 3-hydroxy-3-methylglutaryl-CoA reductase, a key enzyme in plant isoprenoid biosynthesis

LI Rong, WANG Zhe-zhi

(College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China)

Abstract: In the present study, 3-hydroxy-3-methylglutaryl-CoA reductase from rubber, tobacco, capsicum and andrographis, which were registered in GenBank, were analyzed and predicted by the tools of bioinformatics in the following aspects: the composition of nuclei acid sequences and amino acid sequences, signal peptide, transmembrane topological structure, hydrophobicity or hydrophilicity, secondary and tertiary structure of protein, molecular phylogenetic evolution and so on. The results are as follows: the full-length gene of HMGR contains an opening reading frame, 5'-untranstrated region and 3'-untranstrated region; HMGR is a hydrophilic and transmembrane protein which lack of signal peptide; the amino acid sequences of HMGR include two functional HMG-CoA binding motifs and two functional NADPH binding motifs; the main motif of predicted secondary structure of HMGR are alpha helix and random coil, beta turn and extended strand are spreaded in the whole secondary structure of protein; the predicted tertiary structure of functional domain of HMGR includes the helical N-domain, S-domain and L-domain.

Key words: plant isoprenoid biosynthesis; 3-hydroxy-3-methylglutaryl-CoA reductase(HMGR); bioinformatics

收稿日期: 2006-04-10 修回日期: 2006-05-15

基金项目: 国家“十五”科技攻关项目(2004BA721A35)[Supported by Key Technologies Research and Development Program of State Tenth Five-Year Plan Project(2004BA721A35)]

作者简介: 李嵘, 男, 主要从事植物分子生物学方面的研究, (E-mail) lirong@snnu.edu.cn.

萜类化合物(terpenoid)是一类由异戊二烯(isoprene)为结构组成单元的天然烃类化合物的统称,在自然界中分布广泛、种类繁多、结构多样,迄今为止,约有 22 000 种萜及萜类衍生物的化学结构已被鉴定(Sacchettini 等,1997; Verpoorte 等,2000; Lange 等,2000)。据估计,植物体内的萜类化合物有 30 000 种之多(Page 等,2004),它们不仅在植物生命活动中起重要作用,而且被广泛应用于工业、医药卫生等领域。据文献报道,植物体内萜类化合物的合成有甲羟戊酸(mevalonate pathway, MVA)和丙酮酸/磷酸甘油醛(pyruvate/glyceraldehydes-3-phosphate pathway, DXP)两条合成途径,两者都以异戊烯基焦磷酸(isopentenyl diphosphate, IPP)为主要中间产物(Lange 等,1998; Hartmul, 1999; Rohmer, 1999; 陈大华等,2000; Dudareva 等,2005; Liu 等,2005)。3-羟基-3-甲基戊二酰辅酶 A 还原酶在 MVA 途径中催化 3-羟基-3-甲基戊二酰辅酶 A 转化为甲羧戊酸,由于该过程不可逆,故 HMGR 被视作 MVA 途径中的第一个限速酶,是细胞质萜类化合物代谢中的重要调控位点(Bach, 1986; Choi 等,1992)。该酶的活性先后在豇豆(*Vigna vexillata*)、菊芋(*Helianthus tuberosus*)、胡萝卜(*Daucus carota*)、假荆芥(*Nepeta cataria*)等植物中检测到,1986 年被分离提纯,证明是一类与膜结合的亲水蛋白质(Bach, 1986)。目前 HMGR 酶基因已从巴西橡胶树、马铃薯、高地棉、番茄、曼地亚红豆杉等植物中分离克隆(Chye 等,1992; Korth 等,1997; Yang 等,1991; Joost 等,1995; Loguercio 等,1999; Rodriguez-concepcion 等,1999; Kato-emori 等,2001; Liao 等,2004)。生物信息学(Bioinformatics)是一门由分子生物学和信息科学技术相结合的交叉学科。由于人类基因组计划的完成,相关生物学的发展也已从基因组测序转向对基因组表达的分析,对蛋白质结构与功能的预测。利用国际互联网上的各类分子生物学数据库及分析软件对已知的核酸和蛋白质序列进行比对、分析、建立计算模型,从而对其结构、功能进行推断及预测,是当今生物信息学发展的必然潮流。本文利用生物信息学的方法,以橡胶(*Hevea brasiliensis*)为重点,对烟草(*Nicotiana tabacum*)、辣椒(*Capsicum annuum*)、穿心莲(*Andrographis paniculata*)等植物的 3-羟基-3-甲基戊二酰辅酶 A 还原酶的基因及相应氨基酸序列的理化性质、结构特征、功能及系统演化关系等进行预测

和分析,以期在今后深入开展该类酶的相关研究提供一定的理论依据。

1 材料和方法

数据资料来源于 National Center for Biotechnology Information(NCBI)核酸及蛋白质数据库中已注册的 3-羟基-3-甲基戊二酰辅酶 A 还原酶的核酸序列及其对应的氨基酸序列:橡胶(Accession: AY706757, Accession: AAU08214)、烟草(Accession: NTU60452, Accession: AAB87727)、辣椒(Accession: AF110383, Accession: AAD28179)、穿心莲(Accession: AY254389, Accession: AAP14352)。

依据 Vector NTI Suite 8.0 软件及 <http://www.ncbi.nlm.nih.gov/>、<http://www.ebi.ac.uk/>、<http://www.cbs.dtu.dk/>、<http://cn.expasy.org/>等网站提供的各类生物信息学软件进行在线分析(李衍达等,2000; 钟扬等,2003)。核酸及氨基酸序列的组成成分分析、理化性质分析、开放阅读框(Open reading frame, ORF)的查找和翻译利用 Vector NTI Suite 8.0 软件及 ProtParam、pI/MV、ORF Finder 在线工具上进行;核酸及氨基酸序列的同源性比对、多序列比对及分子系统进化关系利用 Blast 和 Clustal W 在线工具完成;蛋白质信号肽的预测、跨膜结构域及亲水性/疏水性的分析利用在线工具 SignalP、TMHMM、ProtScale 完成;蛋白质二级及三级结构的预测利用 SOPMA 和 SWISS-MODEL 在线工具完成。

2 结果和分析

2.1 不同植物 HMGR 基因的核酸及对应氨基酸序列的组成成分及理化性质分析

用 ORF Finder、Vector NTI Suite 8.0、ProtParam 和 pI/MV 对橡胶、烟草、辣椒、穿心莲 HMGR 基因的核酸序列及氨基酸序列进行分析,结果见表 1。其中,橡胶 HMGR 基因的全长 mRNA 序列及其推导的氨基酸序列如图 1 所示。从表 1 看出,不同植物 3-羟基-3-甲基戊二酰辅酶 A 还原酶(HMGR)基因的全长均包括 5'、3'非翻译区(Untranslated region, UTR)和一个开放阅读框,起始密码子均为 atg,终止密码子除辣椒为 taa 外,其余均为 tag,开放阅读框的碱基数及所编码的氨基酸残

基数相差不大。

不同植物 HMGR 的分子量、理论等电点、酸性和碱性氨基酸的比例、总氨基酸带电荷的比例、极性氨基酸的比例、疏水性 (Hydrophobicity) 氨基酸的比例基本一致;含量最丰富的氨基酸均为 Ala、Leu、Ser、Val、Gly;除橡胶 HMGR 属于稳定类蛋白质外,其余均属不稳定类蛋白质。

2.2 不同植物 HMGR 的核酸及氨基酸序列比对分析

用 Blast 程序对橡胶 HMGR 的核酸及氨基酸序列进行同源性比对 (Altschul 等, 1997), 结果表明橡胶 HMGR 的核酸序列与其它植物的核酸序列如甜瓜 *Cucumis melo* (Accession: AB021862)、拟南芥 *Arabidopsis thaliana* (Accession: AY488113)、烟草 (Accession: U60452)、高地棉 *Gossypium hirsutum* (Accession: AF038045)、马铃薯 *Solanum tuberos-*

um (Accession: AB022690)、番茄 *Lycopersicon esculentum* (Accession: M63642) 等具有较高的同源性;但相对于蛋白质-蛋白质的 Blastp 比对来说,橡胶 HMGR 的氨基酸序列与其它植物的氨基酸序列有着更广泛和更高的相似性,如与甜瓜 (Accession: BAA36291)、拟南芥 (Accession: AAR83122)、烟草 (Accession: AAB87727)、高地棉 (Accession: AAC05088)、马铃薯 (Accession: BAA93631)、番茄 (Accession: AAA34169) 的一致性分别为 81%、79%、77%、76%、76%、75%, 相似性分别为 88%、89%、88%、85%、87%、86%。对烟草、辣椒、穿心莲 HMGR 的核酸及氨基酸序列进行类似同源性比对,也得到上述规律,即不同植物氨基酸序列之间的比对相对于核酸序列之间的比对来说,有着更高和更广泛的相似性。

表 1 不同植物 HMGR 基因的核酸序列及对应氨基酸序列的组成成分及理化性质分析
Table 1 A comparison of composition and physical and chemical characters of nucleic acid sequence and deduced amino acid sequence of HMGRs among different plants

项目 Items	橡胶 <i>Hevea brasiliensis</i>	烟草 <i>Nicotiana tabacum</i>	辣椒 <i>Capsicum annuum</i>	穿心莲 <i>Andrographis paniculata</i>
基因全长 Full-length (bp)	1 838	2 473	2 402	2 006
开放阅读框 Open reading frame (bp)	1 728	1 815	1 815	1 671
起始位点及密码子 Site and codon of star (bp)	43, ATG	70, ATG	97, ATG	85, ATG
终止位点及密码子 Site and codon of stop (bp)	1 770, TAG	1 884, TAG	1 911, TAA	1 755, TAG
5'/3'非翻译区长度 UTR of 5'/3' (bp)	42/68	69/589	96/491	84/251
推导氨基酸残基数 Number of amino acid	575 aa	604 aa	604 aa	556 aa
分子量 Molecular weight (kDa)	61.75	64.73	64.88	59.74
理论等电点 pI Theoretical isoelectric point	6.61	6.64	5.69	8.13
含量最丰富的氨基酸 Rich amino acid	Ala (9.91%) Leu (9.74%) Ser (9.39%) Val (8.17%) Gly (7.48%)	Ser (9.93%) Ala (9.77%) Val (9.44%) Leu (9.11%) Gly (7.45%)	Ala (10.43%) Leu (9.11%) Ser (8.28%) Val (7.95%) Gly (7.78%)	Leu (10.25%) Ser (9.17%) Ala (8.99%) Val (8.27%) Gly (7.73%)
酸性/碱性氨基酸比例 Acidic/basic amino acid (%)	11.1/12.1	11.04/12.22	11.99/11.77	10.37/12.65
带电氨基酸比例 Charged amino acid (%)	31	29.68	30.74	30.05
极性氨基酸比例 Polar amino acid (%)	25	26.22	26.27	24.71
疏水性氨基酸比例 Hydrophobicity amino acid (%)	37	36.50	36.37	37.71
蛋白质不稳定性指数 Instability Index of protein	36.83%, 属于稳定类蛋白	46.24%, 属于不稳定类蛋白	40.95%, 属于不稳定类蛋白	43.01%, 属于不稳定类蛋白

用 Clustal W 程序对橡胶、烟草、辣椒、穿心莲、甜瓜、拟南芥、番茄、高地棉、马铃薯、桑树 *Morus alba* (Accession: AAD03789)、苹果 *Malus domestica* (Accession: AAL03986)、曼地亚红豆杉 *Taxus media* (Accession: AAQ82685) 的氨基酸序列进行多序列比对 (Higgins 等, 1994), 结果表明 (图 2), 橡胶、烟草、辣椒、穿心莲 HMGR 的功能域与其它植

物 HMGR 的功能域有着几乎一致的氨基酸组成,也具有 HMGR 活性所必需的典型的多肽位点,即两个 HMG-CoA 结合 motif: EMP(V/I)GY(V/I)QIP 和 TTEGCLVA, 两个 NADPH 结合 motif: DAMGMNM 和 GTVGGGT。其中, TTEGCLVA 中的谷氨酸 Glu 在 HMGR 催化中起着重要作用 (Park 等, 1992; Campos 等, 1995; Loguercio 等,

```

1          cgccgatccccctttctctctcctgcgccgcataatatttac
43 atggacaccaccgcccggctccaccaccgaagcatgctacaccgttgaggaccgttctccgaccactccgaagcgg
M D T T G R L H H R K H A T P V E D R S P T T P K A
121 tcggacgcgcttccgcttccctctacctgaccaacgcgggttttctcacgctgttcttctcggtggcgattacctc
S D A L P L P L Y L T N A V F F T L F F S V A Y Y L
199 cttcaccggtggcgcgacaagatccgcaactccactccccttcataatcgttactctctctgaaattgttctattgtc
L H R W R D K I R N S T P L H I V T L S E I V A I V
277 tcctcattgcctctttcatttaccctcctaggattcttcggatcgtatttggcagtcattcattgacgcgcctcc
S L I A S F I Y L L G F F G I D F V Q S F I A R A S
355 catgacgtgtgggacctcgaagatccgatcccaactacctcatcgatgaagatcaccgtctctgttactgacctccc
H D V W D L E D T D P N Y L I D E D H R L V T C P P
433 gctaataatctactaagactaccattattgcccacactaccaaatgacctacccggaaccccttaattgcaccctta
A N I S T I A A P T K L P T S E P L I A P L
511 gtctcgggaagcgaatgatcgtaactccgtcgtgatgggaagataccctcctattctctggagtcgaagcgc
V S E E D E M I V N S V V D G K I P S Y S L E S K L
589 ggggactgcaaacgagcggctgcgattcgcacgcgaggcttgcagaggatgacaaggaggctcgtggaaggcttgcca
G D C K R A A A I R R E A L Q R M T R R S L E G L P
667 gtagaagggttcgattacgagtcgattttaggacaatgctgtgaaatgcagtgggatacgtgcagattccgggtggg
V E G F D Y E S I L G Q C C E M P V G Y V Q I P V G
745 attgccccggcctgttctgctgaacggggcgggagtactctgttccaatggcaccacggagggttgtttggggcgagc
I A G P L L L N G R E Y S V P M A T T E G C L V A S
823 actaatagagggtgttaaggcgatttacttctcagggtggggccaccagcgttctgttgaaggatggcatgacaagagcg
T N R G C K A I Y L S G G A T S V L L K D G M T R A
901 cctgttgtaagattcgcgtcgcgactagaccgcggagttgaagtcttcttggaggatcgtgacaattttgatacc
P V V R F A S A T R A A E L K F F L E D P D N F D T
979 ttggccgtagtttttaacaagtcagtagatttgcgaggctccaaggcattaatgctcaattgctggtgaagaatctt
L A V V F N K S S R F A R L Q G I K C S I A G K N L
1057 tatataagattcagctacagcactggcgatgcaatggggatgaacatggtttctaaaggggttcaaacgcttcttgaa
Y I R F S Y S T G D A M G M N M V S K G V Q N V L E
1135 tttctcaaagtgatttttctgatatggatgtcattggaatctcaggaaattttgttcggataagaagcctgctgct
F L Q S D F S D M D V I G I S G N F C S D K K P A A
1213 gtaaatggattgaaggacgtggcaaatcagttgtttgtgaggcaattatcaaggaagaggtggtgaagaaggtgtg
V N W I E G R G K S V V C E A I I K E E V V K K V L
1291 aaaaccaatgtggcctccctagctggagcttaacatgctcaagaatcttctgctggtctgctgttgggtgcttgggt
K T N V A S L V E L N M L K N L A G S A V A G A L G G
1369 ggatttaatgcccacgaggaacatcgatctgcaatctttattgccactggccaggatccagcacagaatggttag
G F N A H A G N I V S A I F I A T G Q D P A Q N V E
1447 agttctcattgacattaccatgatggaagctgcaatgatggaagatcctcatatctctgtgaccatgccctccatt
S S H C I T M M E A V N D G K D L H I S V T M P S I
1525 gagtgggtacagtcggaggtggaactcaactgcatctcagtcgtgcttctcaatttgcctgggtgaagggtgca
E V G T V G G G T Q L A S Q S A C L N L L G V K G A
1603 aacaagagtcgcccaggatcaaacactcaaggctccttctgctccatcgtagctggttcagtttggctggtgagctctcc
N K E S P G S N S R L L A A I V A G S V L A G E L S
1681 ttgatgtctccattgacgtggcgacttctcaagagtcacatgaagtacaacagatccagcaaaagatattgtctaaa
L M S A I A A G Q L V K S H M K Y N R S S K D M S K
1759 gctgcatcttagtgggaatctggtcccagcaatgtaaaatgatctaaaataaaatgtggcggagattgtttgaattc
A A S *
1836 cgg
    
```

图 1 橡胶 HMGR 基因的全长 mRNA 序列及推导的氨基酸序列

Fig. 1 The full-length mRNA sequence and deduced amino acid sequence of HMGR in rubber
 起始密码子用方框显示,终止密码子用斜体下划线显示,5'和 3'非翻译区加粗显示,保守结构域用下划线显示。
 The start codon(ATG)was boxed and the stop codon(TAG)was underlined *italically*. The UTR of 5' and 3' were showed in bold fonts. Conserved motif was underlined.

1998;Jain 等,2000;Kato-emori 等,2001;Rupasinghe 等,2001)。这一结果暗示功能结构域在分子进化中具有较高的稳定性,也意味着植物 HMGR 是一个催化甲羟戊酸合成的功能蛋白。

2.3 不同植物 HMGR 信号肽的预测和分析

信号肽(Signal peptide 或 Signal sequence)位于蛋白质 N 端,一般有 16~26 个氨基酸残基,包括

疏水核心区、信号肽的 C 端和 N 端等三部分。蛋白质首先在细胞质基质游离核糖体上起始合成,然后在信号肽引导下,转运至不同细胞器或细胞质基质中的特定部位发挥作用(翟中和等,2000)。因此,信号肽的预测和分析,对正确认识和理解蛋白质的细胞定位及结构域的分选有重要意义。用 SignalP 3.0 Server 对橡胶 HMGR 氨基酸序列的信号肽进

	HMG-CoA binding	HMG-CoA binding
Gossypium hirstum	ESILGQCCEMPVGYEQIPVGIAGPLLLNGREYSVPMATTE	EGCLVASTNRGCKA 277
Malus domestica	ESILGQCCEMPVGYVQIPVGIAGPLMLDGREFSVPMATTE	EGCLVASTNRGFKA 279
Cucumis melo	ESILGQCCEMPVGYVQIPVGIAGPLLLDGFYTVPMATTE	EGCLVASTNRGCKA 279
Arabidopsis thaliana	ESILGQCCEMPVGYIQIPVGIAGPLLLDGYEYSVPMATTE	EGCLVASTNRGCKA 328
Hevea brasiliensis	ESILGQCCEMPVGYVQIPVGIAGPLLLNGREYSVPMATTE	EGCLVASTNRGCKA 267
Morus alba	ESILGQCCEMPVGYVQIPVGIAGPLLLDEFEYSVPMATTE	EGCLVASTNRGCKA 240
Solanum tuberosum	ESILGQCCEMPIGYIQIPVGIAGPLLLNGKEFSVPMATTE	EGCLVASTNRGCKA 269
Lycopersicon esculentum	ESILGQCCEMPIGYVQIPVGIAGPLLLNGKEFSVPMATTE	EGCLVASTNRGCKA 294
Nicotiana tabacum	ESILGQCCEMPVGYVQIPVGIAGPLLLDGREYSVPMATTE	EGCLVASTNRGCKA 296
Capsicum annuum	ESILGQCCEMTIGYVQIPVGIAGPLLLNGREYSVPMATTE	EGCLVASTNRGCKA 296
Andrographis paniculata	DSILGQCCEMPVGYVQIPVGIAGPLLLNGCEYSVPMATTE	EGCLVASTNRGCKA 250
Taxus media	QSI LGQCCEMPIGYVQIPVGVAGPLLLNGFEYMPVATTE	EGCLVASTNSGCKA 281
	:*****:** *****:**: * : ** :***** * **	
	NADPH binding	NADPH binding
Gossypium hirstu	LYLRFSCFTGDAMGMNMVSKGVQNV 357	SIEVGTVGGGTQLASQSACL 547
Malus domestica	LYMRFCSTGDAMGMNMVSKGVQNV 359	SIEVGTVGGGTQLASQSACL 549
Cucumis melo	LYVRFCCSTGDAMGMNMVSKGVQNV 359	SIEVGTVGGGTQLASQSACL 549
Arabidopsis thaliana	AYVRFCCSTGDAMGMNMVSKGVQNV 408	SIEVGTVGGGTQLASQSACL 598
Hevea brasiliensis	LYIRFSYSTGDAMGMNMVSKGVQNV 347	SIEVGTVGGGTQLASQSACL 537
Morus alba	LYMRFCSTGDAMGMNMVSKGVQNA 320	SIEVGTVGGGTQLASQSACL 520
Solanum tuberosum	LYMRFCSTGDAMGMNMVSKGVQNV 349	SIEVGTVGGGTQLASQSACL 539
Lycopersicon esculentum	LYMRLCCSTGDAMGMNMVSKGVQNV 374	SIEVGTVGGGTQLASQSACL 564
Nicotiana tabacum	LYMRFCSTGDAMGMNMVSKGVQNV 391	SIEVGTVGGGTQLASQSACL 541
Capsicum annuum	LHMRFCSTGDAMGMNMVSKGVQNV 391	SIEVGTVGGGTQLASQSACL 511
Andrographis paniculata	LYLRFCCSTGDAMGMNMVSKGVQNV 345	SIEVGTVGGGTQLASQSACL 495
Taxus media	LYMRFCFTGDAMGMNMVSKGVQNV 361	CIEVGTVGGGTHLASQAACL 551
	* : * : *****	. ***** : *****

图 2 不同植物 HMGR 氨基酸序列多重比对

Fig. 2 Multiple alignment of amino acid sequence among different plant HMGRs

“*”, “,”, “.” 分别表示氨基酸完全相同、保守和半保守三种情况。重要功能结构域 motif 用阴影表示, 即两个 HMG-CoA 结合 motif, EMP(V/D)GY(V/I)QIP 和 TTEGCLVA, 两个 NADPH 结合 motif: DAMGMNM 和 GTVGGGT。在 HMGR 催化中起重要作用的谷氨酸 Glu 用方框显示。

The completely identical, conservative and semi-conservative amino acid residues were indicated with “*”, “,” and “.” respectively. The substrate-binding motifs were gray-shaded. HMG-CoA binding motif include EMP(V/D)GY(V/I)QIP and TTEGCLVA; NADPH binding motif include DAMGMNM and GTVGGGT. The glutamic acid residue, essential for catalysis, was boxed.

行预测(Nielsen 等, 1997; Bendtsen 等, 2004), 结果如图 3, 第 30 号氨基酸残基具有最高的原始剪切位点分值 0.186, 第 36 号氨基酸残基具有最高的信号肽分值 0.689, 第 50 号氨基酸残基具有最高的综合剪切位点分值 0.104, 由此看出, 橡胶 HMGR 不存在信号肽酶切位点, 不具有信号肽。可以推断, 橡胶 HMGR 在细胞质中合成后, 不进行蛋白转运, 而保留在细胞质基质中, 直接与代谢底物相作用。

对烟草、辣椒、穿心莲 HMGR 氨基酸序列的信号肽进行分析, 也得到与橡胶一致的结果, 即 HMGR 不存在信号肽酶切位点, 不具有信号肽。可以推断, 植物 HMGR 在细胞质中合成后, 不进行蛋白转运, 而保留在细胞质基质中, 直接与生成甲羟戊酸相关的代谢底物相作用。由于 HMGR 是 MVA 途径中的一个限速酶(Bach, 1986; Choi 等, 1992),

而 MVA 途径仅存在于细胞质中(Liu 等, 2005), 由此看出, 这一推断与萜类代谢 MVA 途径在细胞中的作用部位是相吻合的。

2.4 不同植物 HMGR 跨膜结构域的预测和分析

跨膜结构域是膜内在蛋白与膜脂相结合的主要部位, 一般由 20 个疏水氨基酸残基组成, 形成 α 螺旋, 它固着于细胞膜上起“锚定”作用(翟中和等, 2000)。跨膜结构域的预测和分析, 对正确认识和理解蛋白质的功能、结构、分类、方位及细胞中的作用部位均有着重要的指示意义。用 TMHMM 2.0 Server 对橡胶 HMGR 氨基酸序列的跨膜结构域进行预测, 结果如图 4, 橡胶 HMGR 有两个跨膜区域, 一个位于多肽链第 31 位的脯氨酸 Pro 与第 53 位的亮氨酸 Leu 之间, 共计 23 个氨基酸残基, 另一个位于多肽链第 74 位的异亮氨酸 Ile 与第 96 位的缬氨

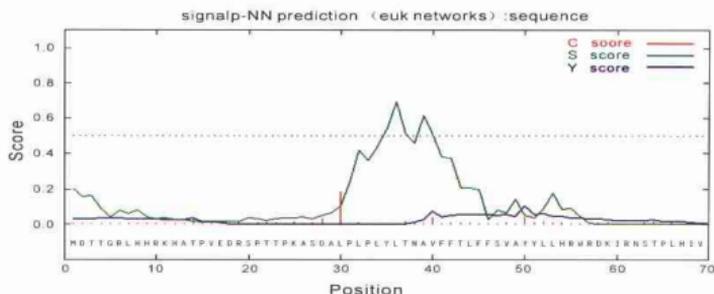


图3 橡胶 HMGR 信号肽预测

Fig. 3 Predicted signal peptide for HMGR in rubber

C score: 原始剪切位点的分值; S score: 信号肽的分值; Y score: 综合剪切位点的分值。

C score: The score for original cleavage site; S score: The score for the signal peptide; Y score: The score for general cleavage sit.

HMGR 在细胞质中合成后, 不进行跨膜转运, 而是直接在细胞质基质中的特定部位行使催化功能。

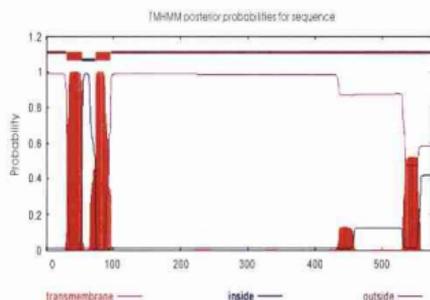


图4 橡胶 HMGR 跨膜结构域的预测

Fig. 4 Predicted transmembrane segments for HMGR in rubber

对烟草、辣椒、穿心莲 HMGR 氨基酸序列的跨膜结构域进行分析, 也得到与橡胶一致的结果, 即 HMGR 有两个跨膜区域, 结合上述信号肽的预测, 可以推断, 植物 HMGR 在细胞质中合成后, 靠其跨膜结构域与细胞膜的相互作用, 将 HMGR “锚定”于细胞质基质中的特定部位行使催化功能。

2.5 不同植物 HMGR 疏水性/亲水性的预测和分析

疏水性是 20 种氨基酸都固有的特性, 即氨基酸远离周围水分子, 将自己包埋进蛋白质核心的相对趋势。这一趋势加上空间立体条件和其它一些因素, 决定一个蛋白质最终折叠成三维空间构象。此

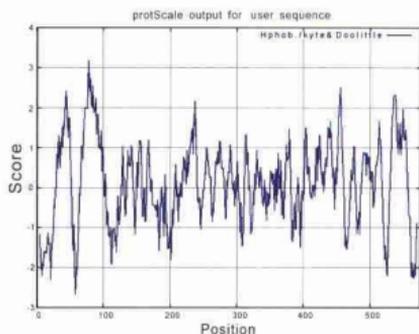


图5 橡胶 HMGR 疏水性/亲水性的预测

Fig. 5 Predicted hydrophobicity/hydrophilicity for HMGR in rubber

外, 通过了解肽链中不同肽段的疏水性, 可对跨膜蛋白的跨膜结构域进行预测。因此, 疏水性/亲水性的预测和分析, 对蛋白质次级结构的预测及功能结构域的分选提供了重要的理论参考。用 ProtScale 对橡胶 HMGR 氨基酸序列的疏水性/亲水性进行预测 (Kyte 等, 1982), 结果如图 5 所示, 多肽链第 58 位的天冬氨酸 Asp 具有最低的分值 -2.667, 亲水性最强; 第 78 位的缬氨酸 Val 具有最高的分值 3.167, 疏水性最强。整个多肽链表现为亲水性, 其中前 100 位的多肽链具有明显的两个疏水区域, 这与前述跨膜结构域的预测完全相符, 说明此段肽链是橡胶 HMGR 的跨膜结构域。

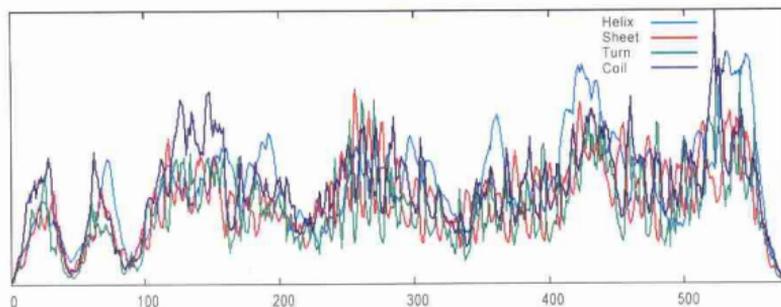


图6 橡胶 HMGR 二级结构的预测

Fig. 6 Predicted secondary structure for HMGR in rubber

Helix; α -螺旋(蓝线); Sheet; 延伸链(红线); Turn; β -转角(绿线); Coil; 不规则盘绕(紫线).

Helix; Alpha helix (blue curve line); Sheet; Extended strand (red curve line); Turn;

Beta turn (green curve line); Coil; Random coil (purple curve line).

对烟草、辣椒、穿心莲 HMGR 氨基酸序列的疏水性/亲水性进行分析,得到与橡胶一致的结果, HMGR 整个多肽链表现为亲水性,在其前 110 位的多肽链具有两个明显的疏水区域,综合前述跨膜结构域的预测结果,可以推断,植物 HMGR 的 N 端具有两个明显的疏水区域,此疏水肽段正是植物 HMGR 的跨膜结构域。

2.6 不同植物 HMGR 二级及三级结构的预测和分析

蛋白质的生物学功能是蛋白质分子的自然构象所具有的性质,由其高级结构所决定,高级结构是由一级结构即氨基酸顺序决定,而氨基酸顺序是由遗传物质 DNA 的碱基顺序所规定。氨基酸的多肽链借助氢键排列成沿一维方向有规则的重复构象的二级结构,它是氨基酸顺序与三维构象之间的桥梁。二级结构借助范德华力、氢键、静电和疏水等相互作用形成蛋白质的三级结构,从而发挥正常的生物学功能(王镜岩等,2002)。因此,蛋白质高级结构的预测和分析,对理解蛋白质结构与功能之间的相关性有着极其重要的意义。用 SOPMA 对橡胶 HMGR 氨基酸序列的二维结构进行预测(Geourjon 等,1995),结果如图 6 所示,橡胶 HMGR 由 41.22% 的 α -螺旋、15.65% 的延伸链、5.74% 的 β -转角和 37.39% 的不规则盘绕组成,纵观蛋白的整体结构, α -螺旋和不规则盘绕是橡胶 HMGR 最大量的结构元件,而 β -转角和延伸链则散布于整个蛋白质中。

对烟草、辣椒、穿心莲 HMGR 氨基酸序列的二维结构进行分析,得到与橡胶一致的结果, α -螺旋和

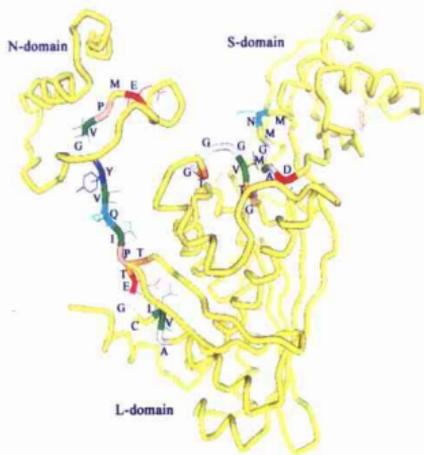


图7 基于同源性建模获得的橡胶 HMGR 的立体结构

Fig. 7 The three-dimensional structure for HMGR in rubber established by homology-based modeling

蓝色英文大写字母示功能结合 motif.

The substrate-binding motifs were in blue and marked with capital letters.

不规则盘绕是 HMGR 整体结构中的主要组成结构元件, β -转角和延伸链散布于整个蛋白质中。由此可以推断,植物 HMGR 整体蛋白结构中, α -螺旋和

不规则盘绕是 HMGR 整体结构中的主要组成结构元件, β -转角和延伸链散布于整个蛋白质中。由此可以推断, 植物 HMGR 整体蛋白结构中, α -螺旋和不规则盘绕是最大量的结构元件, β -转角和延伸链散布于整个蛋白质中。

用 SWISS-MODEL 对橡胶 HMGR 氨基酸序列进行蛋白质三维结构同源性建模 (molecular homologous modeling) (Peitsch, 1995; Guex 等, 1997; Schwede 等, 2003), 结果如图 7 所示, 橡胶 HMGR 的功能区域在空间布局上折叠成“V”形, 螺旋状的 N 结构域和 S 结构域构成“V”形的两臂, S 结构域内藏有一个功能 NADPH 结合 motif (DAMGM-NM), “V”形的中间部分 L 结构域内隐藏着两个功能 HMG-CoA 结合 motif (EMPVGYVQIP 和 TTEGCLVA) 和一个功能 NADPH 结合 motif (GTVGGGT), 这一结果与人类 HMGR 的三维结构非常相似 (Istvan 等, 2000)。

对烟草、辣椒、穿心莲 HMGR 氨基酸序列进行蛋白质三维结构同源性建模, 得到与橡胶一致的结果, 即 HMGR 的功能区域在空间结构上折叠成“V”形, 螺旋状的 N 结构域和 S 结构域组成“V”形的两臂, 中部部分由 L 结构域组成, L 结构域内藏有两个功能 HMG-CoA 结合 motif (EMPVGYVQIP 和 TTEGCLVA) 及一个功能 NADPH 结合 motif (GTVGGGT), S 结构域内藏有一个功能 NADPH 结合 motif (DAMGMNM)。由此可以推断, 植物 HMGR 是一个在甲羟戊酸萜类生物合成途径中起作用的功能蛋白。

2.7 不同植物 HMGR 氨基酸序列的分子系统进化分析

用 Clustal W 程序对橡胶 (Accession: AAU08214)、苹果 (Accession: AAL03986)、番茄 (Accession: AAA34169)、高地棉 (Accession: AAC05088)、马铃薯 (Accession: BAA93631)、甜瓜 (Accession: BAA36291)、拟南芥 (Accession: ARR83122)、桑树 (Accession: AAD03789)、曼地亚红豆杉 (Accession: AAQ82685) 的氨基酸序列进行分子系统进化分析 (Higgins 等, 1994), 结果如图 8, 九种不同植物的 HMGR 氨基酸序列聚成三支, 甜瓜与拟南芥、橡胶、桑树聚成一支; 高地棉与苹果、曼地亚红豆杉聚成第二支; 番茄与马铃薯聚成第三支。这一结果虽与其他学者的研究不一致 (Takhtajan, 1997; Thorne, 2000; Soltis 等, 2000; APG II, 2003;

吴征镒等, 2003, 2004), 但还是反应了九种不同植物间的亲缘关系。如番茄与马铃薯同属茄科 Solanaceae, 亲缘关系最近, 聚为一支。高地棉 (锦葵科 Malvaceae) 与苹果 (蔷薇科 Rosaceae) 虽属不同的科, 但它们间的亲缘关系较为接近, 也聚为一支。与此类似, 甜瓜与拟南芥、橡胶、桑树均依它们间的亲缘关系聚成一支。由此看出, 依据氨基酸序列所得出的系统演化关系虽不能真实反映植物在漫长历史长河中的自然进化关系, 但其结果对判断不同植物间的亲缘关系仍具一定的借鉴意义。

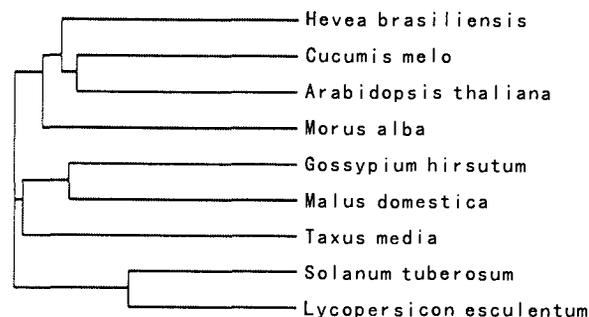


图 8 不同植物 HMGR 的氨基酸序列分子系统发生树
Fig. 8 The phylogenetic tree of amino acid sequence of HMGRs among different plants

3 结论和讨论

随着人类基因组计划的完成及各类模式生物基因组 DNA 大规模的测序, 二十一世纪生命科学的发展由基因组学时代进入到后基因组学时代——蛋白质组学时代。运用生物信息学的理念、方法、工具对海量的生物学数据进行处理和分析, 对揭示和探索生命的奥秘必将产生深远的影响。生物大分子的功能是由其基本组成成分及空间结构所决定, 也即“结构迎合功能”。从生物信息学的角度对核酸及氨基酸序列进行分析, 并对其结构和功能进行预测, 以为生命现象的本质提供一定的理论参考依据。基于此, 本研究结论如下:

(1) 不同植物 HMGR 基因的全长均包括 5'、3' 非翻译区和一个开放阅读框, 开放阅读框的碱基数及所编码的氨基酸残基数相差不大。(2) 不同植物 HMGR 的分子量、理论等电点、酸性和碱性氨基酸的比例、总氨基酸带电荷的比例、极性氨基酸的比例、疏水性氨基酸的比例基本一致; 含量最丰富的氨基酸均为 Ala、Leu、Ser、Val、Gly。(3) 不同植物

HMGR 的核酸序列之间具有较高的同源性;但相对于蛋白质-蛋白质的 Blastp 比对来说,不同植物 HMGR 的氨基酸序列之间有着更广泛和更高的相似性。(4)不同植物 HMGR 的功能域有着几乎一致的氨基酸组成,具有 HMGR 活性所必需的典型的多肽位点,即两个 HMG-CoA 结合 motif: EMP(V/D)GY(V/D)QIP 和 TTEGCLVA,两个 NADPH 结合 motif: DAMGMNM 和 GTVGGGT。其中, TTEGCLVA 中的谷氨酸 Glu 在 HMGR 催化中起着重要作用。(5)植物 HMGR 不存在信号肽酶切位点,不具有信号肽,不进行蛋白转运。(6)植物 HMGR 具有两个跨膜区域,可以推断, HMGR 在细胞质中合成后,靠其跨膜结构域与细胞膜脂的相互作用,将 HMGR“锚定”于细胞质基质中的特定部位行使催化功能。(7)植物 HMGR 的多肽链表现为亲水性,前 110 位的多肽链具有两个明显的疏水区域。(8)植物 HMGR 由 α -螺旋、延伸链、 β -转角和不规则盘绕组成,纵观蛋白的整体结构, α -螺旋和不规则盘绕是植物 HMGR 最大量的结构元件,而 β -转角和延伸链则散布于整个蛋白质中。(9)植物 HMGR 的功能区域在空间结构上折叠成“V”形,螺旋状的 N 结构域和 S 结构域组成“V”形的两臂,中部部分由 L 结构域组成,L 结构域内藏有两个功能 HMG-CoA 结合 motif(EMPVGYVQIP 和 TTEGCLVA)及一个功能 NADPH 结合 motif(GTVGGGT),S 结构域内藏有一个功能 NADPH 结合 motif(DAMGMNM)。(10)不同植物 HMGR 氨基酸序列所得出的系统演化关系虽不能真实的反映植物在漫长历史长河中的自然进化关系,但其结果对判断不同植物之间的亲缘关系仍具有一定的借鉴意义。

上述结论虽属推断和预测,但为后期植物 3-羟基-3-甲基戊二酰辅酶 A 还原酶的功能验证、原核及真核表达载体的构建、RNA 干涉等工作提供了一定的理论参考依据。

参考文献:

- 王镜岩,朱圣庚,徐长法. 2002. 生物化学(上册)[M]. 第三版. 北京:高等教育出版社:157-181.
- 吴征镒,陈心启. 2004. 中国植物志(第 1 卷)[M]. 北京:科学出版社:95-573.
- 吴征镒,路安民,汤彦承,等. 2003. 中国被子植物科属综论[M]. 北京:科学出版社:57-1 071.
- 李衍达,孙之荣,等(译). 2000. Baxevanis AD, Ouellette BFF (著). 生物信息学[M]. 北京:清华大学出版社:231-250.
- 钟扬,王莉,张亮(主译). 2003. Mount DW(著). 生物信息学[M]. 北京:高等教育出版社:301-345.
- 翟中和,王喜忠,丁明孝. 2000. 细胞生物学[M]. 北京:高等教育出版社:78-193.
- Altschul S F, Madden T L, Schaffer A A, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs[J]. *Nucleic Acids Res*, **25**(17): 3 389-3 402.
- ANGIOSPERM PHYLOGENY GROUP(APG II). 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II [J]. *Botanical Journal of the Linnean Society*, **141**: 399-436.
- Bach T J. 1986. Hydroxymethylglutaryl-CoA reductase, a key enzyme in phytosterol synthesis[J]. *Lipids*, **21**(1): 82-88.
- Bendtsen J D, Nielsen H, Heijne G V, et al. 2004. Prediction of signal peptides: SignalP 3. 0 [J]. *J Mol Biol*, **340**: 783-795.
- Campos N, Boronat A. 1995. Targeting and topology in the membrane of plant 3-hydroxy-3-methylglutaryl coenzyme A reductase[J]. *Plant Cell*, **7**(12): 2 163-2 174.
- Chen DH(陈大华), Ye HC(叶和春), Li GF(李国风), et al. 2000. Advances in molecular biology of plant isoprenoid metabolic pathway (植物类异戊二烯代谢途径的分子生物学研究进展)[J]. *Acta Bot Sin* (植物学报), **42**(6): 551-558.
- Choi D, Ward B L, Bostock R M. 1992. Differential induction and suppression of potato 3-hydroxy-3-methylglutaryl Coenzyme A reductase genes in response to phytophthora infestans and to its elicitor arachidonic acid[J]. *The Plant Cell*, **4**(10): 1 333-1 344.
- Chye M L, Tan C T, Chua N H. 1992. Three genes encode 3-hydroxy-3-methylglutaryl-CoA reductase in *Hevea brasiliensis*: hmg1 and hmg3 are differentially expressed[J]. *Plant Mol Biol*, **19**(3): 473-484.
- Dudareva N, Andersson S, Orlova I, et al. 2005. The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers[J]. *Proc Natl Acad Sci USA*, **102**(3): 933-938.
- Geourjon C, Deléage G. 1995. SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments[J]. *Comput Appl Biosci*, **11**(6): 681-684.
- Guex N, Peitsch M C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling[J]. *Electrophoresis*, **18**: 2 714-2 723.
- Hartmul K L. 1999. The 1-deoxyxylulose-5-phosphate pathway of isoprenoid biosynthesis in plant[J]. *Annu Rev Plant Physiol Plant Mol Biol*, **50**: 47-65.
- Higgins D, Thompson J, Gibson T, et al. 1994. CLUSTAL W; improving the sensitivity of progressive multiple sequence

- alignment through sequence weighting, position-specific gap penalties and weight matrix choice[J]. *Nucleic Acids Res*, **22**;4 673—4 680.
- Istvan E S, Deisenhofer J. 2000. The structure of the catalytic portion of human HMG-CoA reductase[J]. *Archives of Biochemistry and Biophysics*, **1** 529;9—18.
- Jain A K, Vincent R M, Nessler C L. 2000. Molecular characterization of a hydroxyl methylglutaryl -CoA reductase gene from mulberry (*Morus alba* L.) [J]. *Plant Mol Biol*, **42** (4), 559—569.
- Joost O, Bianchini G, Bell A A, *et al.* 1995. Differential induction of 3-hydroxy-3-methylglutaryl CoA reductase in two cotton species following inoculation with *Verticillium* [J]. *Mol Plant Microbe Interact*, **8**(6); 880—885.
- Kato-emori S, Higashi K, Hosoya K, *et al.* 2001. Cloning and characterization of the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase in melon (*Cucumis melo* L. *reticulatus*) [J]. *Mol Genet Genomics*, **265**(1): 135—142.
- Korth K L, Stermer B A, Bhattacharyya M K, *et al.* 1997. HMG-CoA reductase gene families that differentially accumulate transcripts in potato tubers are developmentally expressed in floral tissues [J]. *Plant Mol Biol*, **33**(3): 545—551.
- Kyte J, Doolittle R F. 1982. A simple method for displaying the hydropathic character of a protein [J]. *J Mol Biol*, **157** (6): 105—132.
- Lange B M, Rujan T, Martin W, *et al.* 2000. Isoprenoid biosynthesis; the evolution of two ancient and distant pathways across genomes [J]. *Proc Natl Acad Sci USA*, **97**(24): 13 172—13 177.
- Lange B M, Wilding M R, Mccaskill D, *et al.* 1998. A family of transketolases that directs isoprenoid biosynthesis via a mevalonate independent pathway [J]. *Proc Natl Acad Sci USA*, **95**;2 100—2 104.
- Liao ZH, Tan QM, Chai YR, *et al.* 2004. Cloning and characterization of the gene encoding HMG-CoA reductase from *Taxus media* and its functional identification in yeast [J]. *Functional Plant Bio*, **31**(1); 73—81.
- Liu Y, Wang H, Ye HC, *et al.* 2005. Advances in the plant isoprenoid biosynthesis pathway and its metabolic engineering [J]. *J Integrative Plant Bio*, **47**(7): 769—782.
- Loguercio LL, Scott HC, Trolinder NL, *et al.* 1999. Hmg-coA reductase gene family in cotton (*Gossypium hirsutum* L.); unique structural features and differential expression of hmg2 potentially associated with synthesis of specific isoprenoids in developing embryos [J]. *Plant Cell Physiol*, **40**(7): 750—761.
- Loguercio L L, Wilkins T A. 1998. Two genomic clones encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase from cotton (*Gossypium hirsutum* L.) [J]. *Plant Physiol*, **116** (2); 867—869.
- Nielsen H, Engelbrecht J, Brunak S, *et al.* 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites [J]. *Protein Engineering*, **10**(1): 1—6.
- Page J E, Hause G, Raschke M, *et al.* 2004. Functional analysis of the final steps of the 1-deoxy-D-xylulose-5-phosphate (DXP) pathway to isoprenoids in plants using virus-induced gene silencing [J]. *Plant Physiol*, **134**: 1 401—1 413.
- Park H, Denbow C J, Cramer C L. 1992. Structure and nucleotide sequence of tomato HMG2 encoding 3-hydroxy-3-methyl-glutaryl coenzyme A reductase [J]. *Plant Mol Biol*, **20**(2); 327—331.
- Peitsch M C. 1995. Protein modeling by E-mail [J]. *Bio/Technology*, **13**; 658—660.
- Rodriguez-concepcion M, Gruissem W. 1999. Arachidonic acid alters tomato HMG expression and fruit growth and induces 3-hydroxy-3-methylglutaryl coenzyme A reductase independent lycopene accumulation [J]. *Plant Physiol*, **119**(1); 41—48.
- Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and high plants [J]. *Nat Prod Rep*, **16**; 565—574.
- Rupasinghe H P V, Almqvist K C, Paliyath G, *et al.* 2001. Cloning of hmg1 and hmg2 cDNAs encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase and their expression and activity in relation to alpha-farnesene synthesis in apple [J]. *Plant Physiol Biochem*, **39**(11); 933—947.
- Sacchettini J C, Poulter C D. 1997. Creating isoprenoid diversity [J]. *Sciences*, **277**(5333); 1 788—1 799.
- Schwede T, Kopp J, Guex N, *et al.* 2003. SWISS-MODEL; an automated protein homology- modeling server [J]. *Nucleic Acids Research*, **31**(13); 3 381—3 385.
- Soltis D E, Soltis P S, Chase M W, *et al.* 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences [J]. *Botanical Journal of the Linnean Society*, **133**; 381—461.
- Takhtajan A. 1997. Diversity and classification of flowering plants [M]. New York: Columbia University Press, 43—265.
- Thorne R F. 2000. The classification and geography of the flowering plants; dicotyledons of the class Angiospermae [J]. *Bot Rev*, **66**(4); 442—664.
- Verpoorte R, Alfermann A W. 2000. Metabolic Engineering of Plant Secondary Metabolism [M]. Dordrecht: Kluwer Academic Publishers, 1—30.
- Yang Z, Park H, Lacy G H, *et al.* 1991. Differential activation of potato 3-hydroxy-3- methylglutaryl coenzyme A reductase genes by wounding and pathogen challenge [J]. *Plant Cell*, **3** (4); 397—405.