

# IPI 基因的生物信息学分析

邢书凡 (10211064)

北京大学生命科学学院,北京,100871,中国

**摘要** 类异戊二烯类物质在生物体的生长发育过程中起着很重要的作用.在植物中这类物质的合成是通过两条途径来完成的.一,定位于细胞质中的 MVA 途径;二,定位于质体中的 MEP 途径.由 IPI 基因所翻译产生的异戊烯焦磷酸酯同分异构酶,在这条途径中的作用还不是很清楚.为此我们对于拟南芥中的 IPI 基因进行分析.

**Keywords:** IPI,类异戊二烯,异戊二烯焦磷酸酯同分异构酶, 生物信息学

**DOI:**自然科学基金

## 一, 背景介绍

### 1.1 IPI 简介

异戊烯焦磷酸同分异构酶(isopentenyl diphosphate isomerase)简称 IPI,在生物体内类异戊二烯类物质的合成中发挥重要作用.其生化酶活可以催化异戊烯焦磷酸酯(IPP)和甲基丙稀基焦磷酸酯(DMAPP)间的同分异构转化.随后,IPP 和 DMAPP 之间,以及 IPP 和 DMAPP 缩合产生的中间产物间,可以发生多次聚合反应,最终合成出各种各样的类异戊二烯类物质.这些类异戊二烯类物质参于生物体内的多种生命活动,如信号传导,生理调节等.

IPI 广泛并保守地存在于生物界中.其中拟南芥全基因组共有两个编码 IPI 的基因.在原先的研究中,Michael C.将他们分别命名为 *IPP1* 和 *IPP2*(Michael C. *et al.* 1998).

### 1.2 IPI 所参与的代谢途径

IPI 参与的总代谢途径共有两条,分别是二羟甲基戊酸合成途径(简称 MVA 途径)和甲基苏糖醇磷酸酯途径(简称 MEP 途径或 Rohmer 途径).

在高等植物细胞中,两途径均存在.其中 MVA 途径定位于细胞质,MEP 途径则定位于质体中.植物体内各种类异戊二烯类物质的合成都是通过以上两条途径进行的.

MVA 途径主要合成的类异戊二烯类物质为固醇,倍半萜,泛醌和多萜等(见图 1).固醇可以参与细胞膜的形成,泛醌可以参与形成线粒体的电子传递链,倍半萜可以修饰特定蛋白,从而有利于蛋白的活性调节以及确定蛋白的亚细胞定位.多萜中的油菜素内酯,是一种很重要的植物激素.

MEV途径主要合成的类异戊二烯类物质,与MVP途径有些差别,包括单萜,特定的倍半萜,二萜,类胡萝卜素,叶绿素和质体醌等(见图1).其中单萜,二萜和特定的倍半萜可以吸引传粉昆虫,有的也可以参与抵御病菌侵染.叶绿素和类胡萝卜素是光合色素.质体醌参与组成光合传递链.MEV途径的合成产物中还有赤霉素(GA),脱落酸(ABA)和细胞分裂素(cytokinin)这些重要的植物激素.

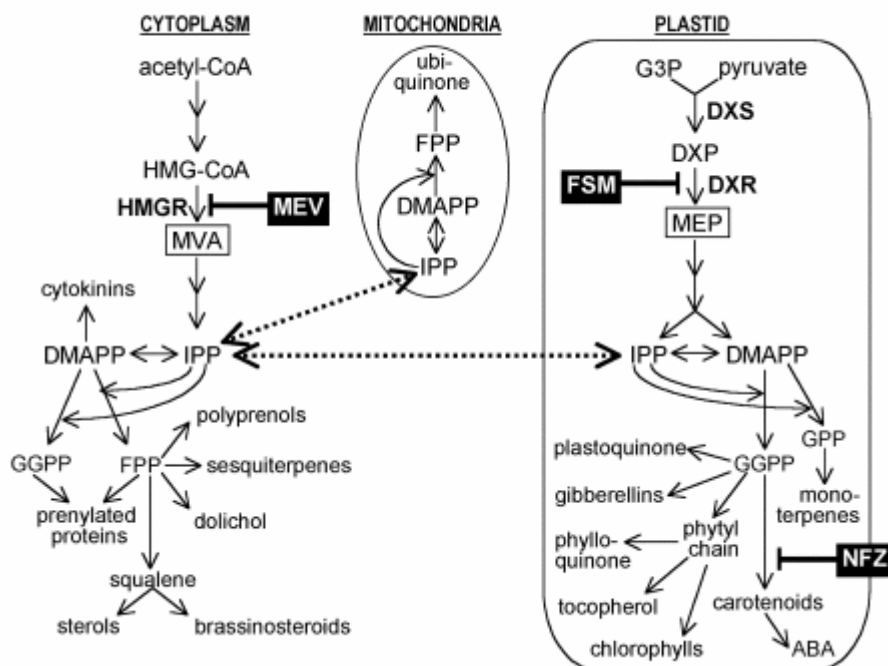


图 1, 植物类异戊二烯类物质合成途径

### 1.3 IPI 研究进展

植物中的 IPI, 最早是 1995 年 Blanc 等在 *Clarkia breweri* 和 *Clarkia xantiana* 中克隆到的. 而真正有功能的 IPI, 是 1998 年 Campbell 等在拟南芥中克隆到的. 接下来 2000, Han Kyung-Hwan 等又在橡胶树 (*Hevea brasiliensis*) 中克隆了 IPI 的编码基因. 随后, 人们又获得了水稻和烟草 (*Nicotiana tabacum*) 中的 IPI 基因克隆.

在对植物 IPI 的研究中, 拟南芥中 IPI 的研究最为深入. 通过体外实验, 人们从原核表达系统收集到具有生化活性的蛋白上清液. 实验证明还拟南芥基因组中共有两个 IPI 编码基因, 基因在基因组中的拷贝数少于二, 同时还预测这两个基因都具有定位于质体的导肽序列. 在 14 天的拟南芥幼苗中根中的 IPI 表达量是叶片的两倍以上.

最近, 对烟草进行的体内 IPI 翻译抑制实验证明, IPI 在植物的类异戊二烯合成途径中起到一定的作用 (Toni M. et al. 2004). 试验中获得了具有黄化斑点的叶片. 通过显微分析表明, 在 IPI 翻译抑制植株中, 叶片的叶绿体在黄化斑点处发育异常. 此外还在实验中发现, IPI 可能有利于 IPP 的稳定存在.

### 1.5 IPI 的研究意义

植物中众多的类异戊二烯类物质的合成都需要有 IPI 这种异构酶的参与. 进一步了解该蛋白的生化活性调控机制, 以及对该蛋白进行改造, 以便提高感控制其生化活性在植物体内的发挥将具有重要意义. 首先在农业生产中有利于农作物的改造. 其次, 该酶在生物界是广泛存在的, 而且参与人体内固醇类激素的合成过程. 因此对该酶的生物活性的深入认识还将对新型药物合成提供线索.

## 二, 拟南芥 IPI 蛋白的生物信息学分析

### 2. 1 拟南芥中 IPI 基因的获得。

首先根据原有的生物学实验数据得知，在拟南芥中共存在两个 *AtIPI* 基因。本实验室克隆了该基因，其测序结果在 NCBI 中做 **Blastn** 比对，其结果如下。

```
□>gi|18086467|gb|AY065053.1 U Arabidopsis thaliana AT5g16440/MQK4_17 mRNA, complete cds
Length = 1037
```

```
Score = 1191 bits (601), Expect = 0.0
Identities = 601/601 (100%)
Strand = Plus / Minus
```

```
Query: 27 ttcagagcttgtgaatggttttcatgtctgcagcttcagtgatagttcctttctcaacat 86
          |||
Sbjct: 873 ttcagagcttgtgaatggttttcatgtctgcagcttcagtgatagttcctttctcaacat 814
```

```
Query: 87 gatccccaccacttcatcaagaaattatccaccaccaatctgaaccatggagatagtttca 146
          |||
Sbjct: 813 gatccccaccacttcatcaagaaattatccaccaccaatctgaaccatggagatagtttca 754
```

```
Query: 147 cagcttcatcgccagcatctgctttcttaccagctccttaagctcttccctgctcagct 206
          |||
Sbjct: 753 cagcttcatcgccagcatctgctttcttaccagctccttaagctcttccctgctcagct 694
```

```
Query: 207 acttgatctcagccacttcatctgggtttggttgaagcttcacatcccgcacgatgaaga 266
          |||
Sbjct: 693 acttgatctcagccacttcatctgggtttggttgaagcttcacatcccgcacgatgaaga 634
```

```
Query: 267 gtatagatgtaacttctgctctccccatttcccacagaaggtgcttgttaaagcatgc 326
          |||
Sbjct: 633 gtatagatgtaacttctgctctccccatttcccacagaaggtgcttgttaaagcatgc 574
```

```
□>gi|22022580|gb|AY127023.1 U Arabidopsis thaliana AT3g02780/F13E7_28 mRNA, complete cds
Length = 1015
```

```
Score = 672 bits (339), Expect = 0.0
Identities = 531/595 (89%)
Strand = Plus / Minus
```

```
Query: 27 ttcagagcttgtgaatggttttcatgtctgcagcttcagtgatagttcctttctcaacat 86
          |||
Sbjct: 875 ttcagagcttgtgaatggttttcatgtctgcagcttcagtgatagttcctttctcaacat 816
```

```
Query: 87 gatccccaccacttcatcaagaaattatccaccaccaatctgaaccatggagatagtttca 146
          |||
Sbjct: 815 gatccccaccacttcatcaagaaattatccaccaccaatctgaaccatggagatagtttca 756
```

```
Query: 147 cagcttcatcgccagcatctgctttcttaccagctccttaagctcttccctgctcagct 206
          |||
Sbjct: 755 aaccttctcaccctgcatctgctttcttaccagctcctttagctcttcccggtcaccat 696
```

```
Query: 207 acttgatctcagccacttcatctgggtttggttgaagcttcacatcccgcacgatgaaga 266
          |||
Sbjct: 695 acttgatctcagctacttcatctgggtttggttgaagcttcacatcccgcacgatgaaga 636
```

由以上结果得知，**IPI** 在拟南芥中的两个同源基因分别是 **At5g16440** 和 **At3g02780**。其蛋白序列分别是。

**AtIPI1**: MTDSNDAGMDAVQRRLMFEDECILVDENDRVVGHDTKYNCHLME

KIEAENLLHRAFSVFLFNSKYELLLQQRSKTKVTFPLVWTNTCCSHPLYRESELEEN  
 VLGVRNAAQRKLFDELGIVAEDVPVDEFTPLGRMLYKAPSDGKWGEHEVDYLLFIVRD  
 VKLQPNPDEVAEIKYVSREELKELVKKADAGDEAVKLSPWFRLLVVDNFLMKWWDHVEK  
 GTITEADMKTIHKL

AtIPI2: MSASSLFLNPLIRLRSLALSSSFSSFRFAHRPLSSISPRKLPNF

RAFSGTAMTDTKDAGMDAVQRRMLFEDECILVDETDREVVGHDSKYNCHLMENIEAKNL  
 LHRAFSVFLFNSKYELLLQQRSNKTKVTFPLVWTNTCCSHPLYRESELIQDNALGVRNA  
 AQRKLLDELGIVAEDVPVDEFTPLGRMLYKAPSDGKWGEHELDYLLFIVRDVKVQPNP  
 DEVAEIKYVSREELKELVKKADAGEEGLKLSPWFRLLVVDNFLMKWWDHVEKGTLVEAI  
 DMKTIHKL

## 2.2 AtIPI 同源蛋白的搜索

在 NCBI 中所找到的 IPI 同源蛋白有 11 个（表 1）。

表 1, IPI 同源基因

1	H.sapiens	IDI1	isopentenyl-diphosphate delta isomerase
2	M.musculus	Idi1	isopentenyl-diphosphate delta isomerase
3	R.norvegicus	Idi1	isopentenyl-diphosphate delta isomerase
4	D.melanogaster	CG5919	Drosophila melanogaster CG5919 gene
5	A.gambiae	1281472	Anopheles gambiae str. PEST ENSANGG0000000...
6	C.elegans	3J183	isopentenyl-diphosphate delta-isomerase (3...
7	S.pombe	idi1	Schizosaccharomyces pombe idi1 gene
8	S.cerevisiae	IDI1	Saccharomyces cerevisiae IDI1 gene
9	N.crassa	NCU07719.1	Neurospora crassa NCU07719.1 gene
10	M.grisea	MG07125.4	Magnaporthe grisea 70-15 MG07125.4 gene
11	A.thaliana	At5g16440	Arabidopsis thaliana At5g16440 gene
12	A.thaliana	At3g02780	Arabidopsis thaliana At3g02780 gene

## 2.3 IPI 同源基因的蛋白序列分析

对各个基因的序列进行结构域分析发现，这些基因均具有保守的 Nudix 结构域。该结构域，在原先的研究中发现具有磷酸水解酶的活性。各基因中结构域的分布如图 2

AtIPI1	.....	0
AtIPI2	.....MSASSLFNLPILIRLRSLLSSFSFRFAHRPLSS	34
Caenorhabdit	.....	0
Homo_sapiens	MWRGLALARAIGCAARGRGQWAVRAADCAQSGRHPGPAVV	40
yeast	...MTADNNSMPHGAVSSYAKLVQNQTPEDILEEFPEII	36
Consensus		
AtIPI1	.....MTDSNDAGMDAVQRRLMFEDECIL	24
AtIPI2	ISPRKLPNFRAFSGTAMDTKDGMDAVQRRLMFEDECIL	74
Caenorhabdit	..MKPSLRGLSQIGMARRAVELSSYDAQQVEYMREQCIN	37
Homo_sapiens	CGRRLISVLEQIRHFVMMPEINTNHLDKQQVQLLAEMCIL	80
yeast	PLQQRPNTRSSETSNDESGETCFSGHDEEQIKLMNENCIV	76
Consensus		ci
AtIPI1	VDENDRVVGHDTKYNCHLMEKIEAENLLHRAFSVFLFNSK	64
AtIPI2	VDETDRVVGHDSKYNCHLMENIEAKNLLHRAFSVFLFNSK	114
Caenorhabdit	VDENDRIIGGVSKREAHSSYQLV...LHRAFSVFSFTSD	73
Homo_sapiens	IDENDNKIGAEIKKNCHLMENIEKGLLHRAFSVFLFNTE	119
yeast	LDWDDNAIGAGTKKVCHLMENIEKGLLHRAFSVFLFNEQG	116
Consensus		d d g k h
AtIPI1	YELLLQORSKTKVTFPLVWNTCCSHPLYRESELI EENVL	104
AtIPI2	YELLLQORSNTKVTFFPLVWNTCCSHPLYRESELIQDNAL	154
Caenorhabdit	NKLLMQKRSAEKITFPNLWNTCCSHPLHTQQEEMDGVPGA	113
Homo_sapiens	NKLLLQCRSDAKITFPGCFTNTCCSHPLSNPAELEESDAL	159
yeast	E...QCRATEKITFPDLWNTCCSHPLCIDDELGLKGKL	152
Consensus		q r k ttf tntccshpl e
AtIPI1	GVRNAAQRKLFD...ELGIVAEVDPVDEF TPLG.RMLY	138
AtIPI2	GVRNAAQRKLLD...ELGIVAEVDPVDEF TPLG.RMLY	188
Caenorhabdit	K..RAAIRKLEHELGITGVSLNKFQ MAGKYTYRAEMNNAP	151
Homo_sapiens	GVRRAAQRRLKA...ELGIPL EEPPEEINYLT.RIHY	193
yeast	DDKIKGAITA AVRKLDHELGIPEDETKTRGKFHFLNR IHY	192
Consensus		
AtIPI1	KAPSDGKWGEHEVDYLLFIVRDVKLQPNPDEVAEIKYVSR	178
AtIPI2	KAPSDGKWGEHELDYLLFIVRDVKVQPNPDEVAEIKYVSR	228
Caenorhabdit	.....WGEHELDYALILRGIGAENCNINENEVAEVREV	184
Homo_sapiens	KAQSDGIWGEHEIDYILLVRKNVTLNPDNEIKSYCYVSK	233
yeast	MAPSNEPWGEHEIDYILFYKINAKENLTVNPNVNEVRDFK	232
Consensus		wgehe dy l
AtIPI1	EELKELVKKADAGDEAV.KLSPWFRLVVDNFI MKWWDHVE	217
AtIPI2	EELKELVKKADAGEEGL.KLSPWFRLVVDNFI MKWWDHVE	267
Caenorhabdit	VLDELKLVWRKEPNSFTPWLKLF SQTETFEKWWTINKNTE	224
Homo_sapiens	EELKELLKKAASGEIKI.T..PWFKIIAATFLFKWWDNLN	270
yeast	WVSPNDLKTMFADPSYKFT..PWFKIIICENYLFNWWEQLD	270
Consensus		
AtIPI1	KGTITEAADMKTIIHK	232
AtIPI2	KGTLVEAIDMKTIIHK	282
Caenorhabdit	ESIEDTNIYKLIH	236
Homo_sapiens	HLNQQFVDHEKIIYRM	284
yeast	DLSEVENDRQIHRML	285
Consensus		

H.sapiens IDI1 isopentenyl-diphosphate delta isomerase.	NP_004499.2 284 aa	
M.musculus Idi1 isopentenyl-diphosphate delta isomerase.	NP_808875.1 227 aa	
R.norvegicus Idi1 isopentenyl-diphosphate delta isomerase.	NP_445991.1 227 aa	
D.melanogaster CG5919 Drosophila melanogaster CG5919 gene.	NP_650962.1 256 aa	
A.gambiae 1281472 Anopheles gambiae str. PEST ENSANGG00000009154 gene.	XP_321388.1 266 aa	
C.elegans 3J183 isopentenyl-diphosphate delta-isomerase (3J183).	NP_498766.1 236 aa	
S.pombe idi1 Schizosaccharomyces pombe idi1 gene.	NP_595164.1 227 aa	
S.cerevisiae IDI1 Saccharomyces cerevisiae IDI1 gene.	NP_015208.1 288 aa	
N.crassa NCU07719.1 Neurospora crassa NCU07719.1 gene.	XP_328425.1 255 aa	
M.grisea MG07125.4 Magnaporthe grisea 70-15 MG07125.4 gene.	XP_367200.1 254 aa	
A.thaliana At5g16440 Arabidopsis thaliana At5g16440 gene.	NP_197148.2 233 aa	
A.thaliana At3g02780 Arabidopsis thaliana At3g02780 gene.	NP_186927.1 284 aa	

图 2 IPI 同源基因蛋白结构示意图

阴影部分表示的是 Nudix 结构域

对以上基因的保守结构域进行建树分析（见图 3）发现，拟南芥中的两个 IPI 基因具有较近的进化距离。同时拟南芥中的 IPI 基因同人类的 IPI 基因具有更高的同源性。

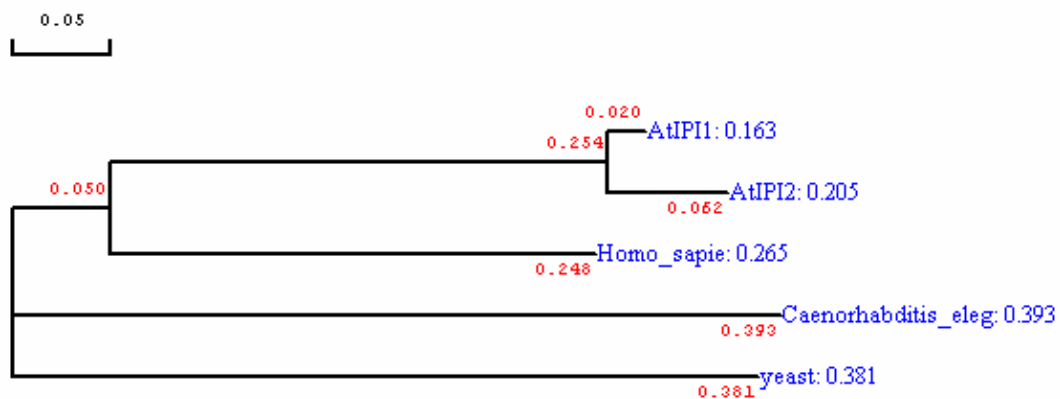
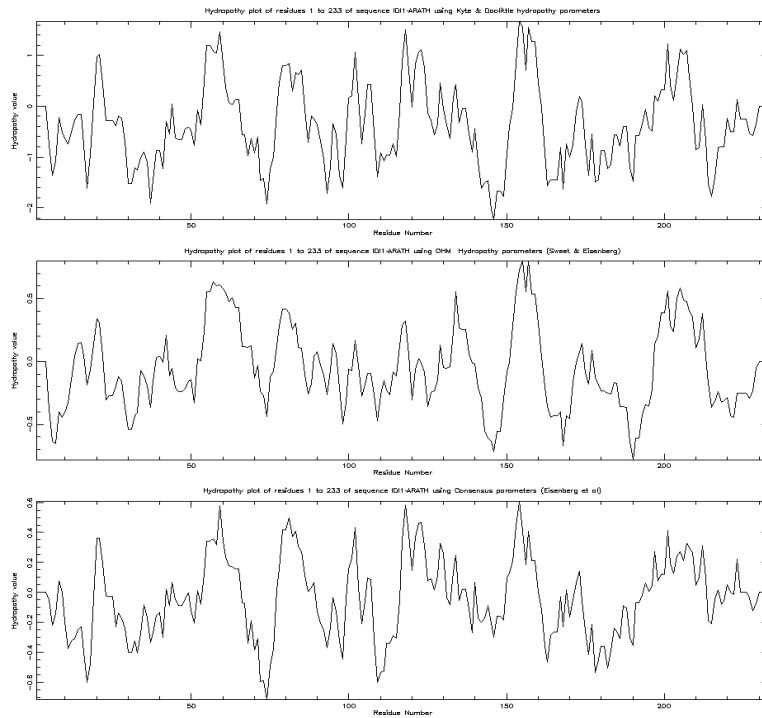


图 3 IPI 基因的建树分析结果

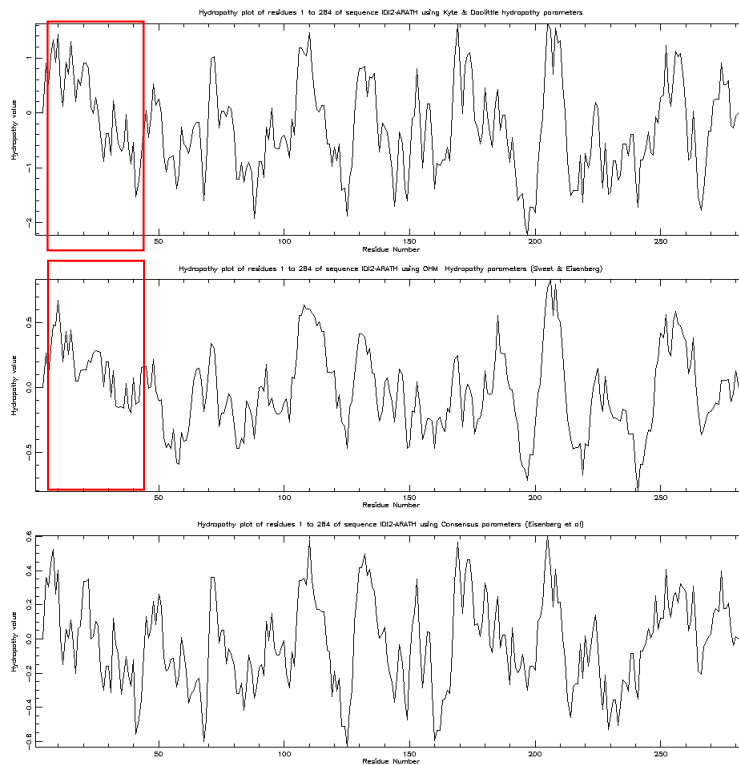
## 2.4 AtIPI 的蛋白结构分析

### 2.4.1 一级结构分析

## AtIPI1:



## AtIPI2



由以上结果比较发现, AtIPI2 在 N 端比 AtIPI1 多出一部分 50 氨基酸的疏水结构域。由 TargetP V1.0 和 Chlorop1.1 的蛋白定位预测结果进一步验证发现, AtIPI2 的前 50 个氨基酸具有叶绿体导肽的作用。

TargetP V1.0 预测结果

	蛋白长度	导肽长度
<b>IPI2</b>	<b>284</b>	<b>45</b>
<b>IPI1</b>	<b>233</b>	-

ChloroP 1.1 预测结果

	蛋白长度	是否有叶绿体定位导肽	导肽长度
<b>IPI2</b>	<b>284</b>	<b>Y</b>	<b>45</b>
<b>IPI1</b>	<b>233</b>	-	-

TargetP V1.0 及 ChloroP 1.1 蛋白定位预测结果

2.4.2 AtIPI 蛋白的三级结构分析

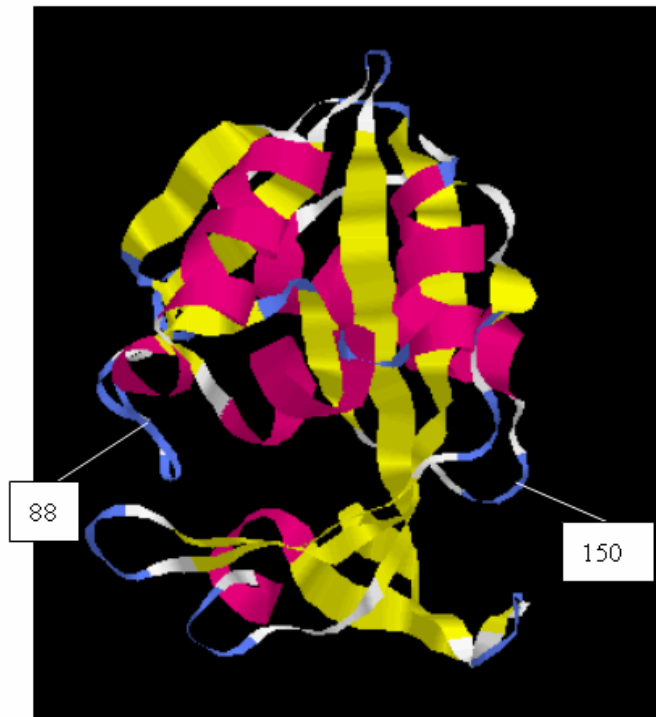


图 4: AtIPI1 的蛋白三级结构预测结果

标识出的氨基酸为该蛋白的，蛋白的预测活性位点

由蛋白三级结构的预测结果（见图 4）可见，在保守的蛋白序列比对中显示的保守结构域部分为分布在蛋白的疏水核心区外的 Coil 区域。最为保守的活性位点所在位置如图所示。其分别是位于蛋白 88 位的半胱氨酸和位于 150 位的谷氨酸。可见，该活性位点有利于蛋白与细胞质中的反应底物充分接触，从而发挥酶的催化作用。

与此同时对 AtIPI1 和 AtIPI2 的蛋白质结构预测结果比较发现。尽管两蛋白在序列的 N 端相差 50 氨基酸的疏水区域，但是在结构上并没有明显区别（见图 5）。

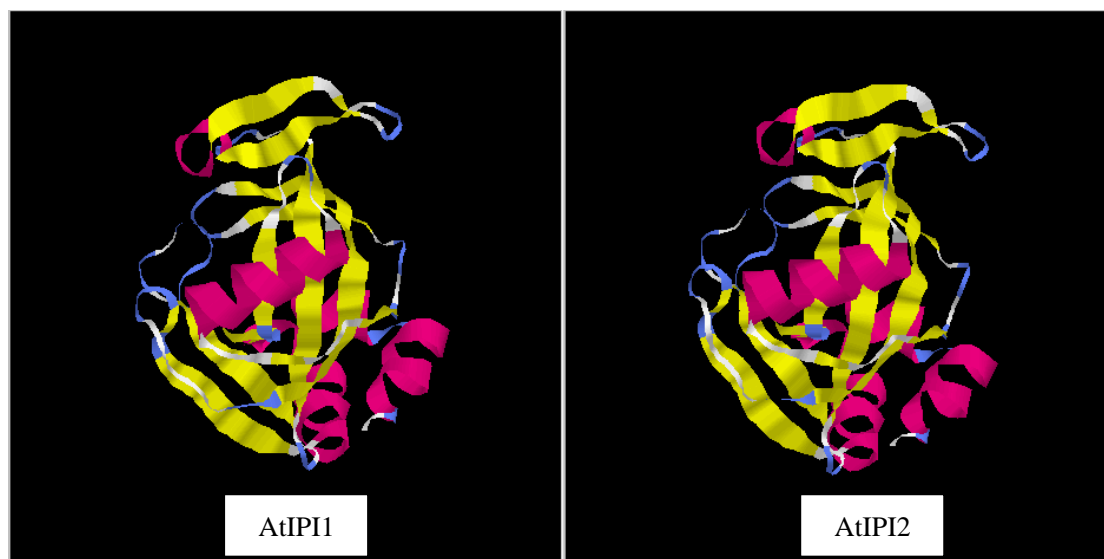


图 5, AtIPI1 和 AtIPI2 蛋白预测三级结构比较

### 三，生物信息学的分析结果对以后生物实验的启发

由以上的分析结果可知 IPI 是广泛并保守地存在于整个生物界中的。现在酵母中的 IPI 研究较为深入，但其他的生物中 IPI 的研究尚未收到重视。因此可以将已有的结果，扩展应用到其他生物的研究中。

在植物中，亚细胞结构有独特的叶绿体组分。由研究表明叶绿体中同样具有类异戊二烯类物质的合成途径。并且由生物信息学的分析也表明 AtIPI2 是定位于叶绿体中的。但是在蛋白结构预测结果中显示，AtIPI1 和 AtIPI2 具有相似的结构。这一结果证明，在以后的生物学试验中可以将 AtIPI1 和 AtIPI2 相互对照进行研究。两蛋白可能具有相似的功能。

并且针对，由生物信息学预测的活性位点，我们还可以将活性位点处的氨基酸突变或替换之后，检测蛋白活性的变化情况。这些结果将对我们以后对蛋白的改造提供线索。

### 四，参考文献

1. Douglas J. McGarvey and Rodney Croteau. (1995) Terpenoid Metabolism. *The Plant Cell*, 7:1015-1026.
2. Eo. Lange. (1998) Molecular biology of gibberellin synthesis. *Planta*, 204:409-419.
3. Francis X. Cunningham And Elisabeth Gantt. (2000) Identification of Multi-Gene Families Encoding Isopentenyl Diphosphate Isomerase in Plants by Heterologous Complementation in *Escherichia coli*. *Plant Cell Physiol.*, 41(1):119-123.
4. Frederick M. Hanhn, Anthony P. Hurlbert, and C. Dale Poulter. (1999) *Escherichia coli* Open Reading Frame 696 Is idi, a Nonessential Gene Encoding Isopentenyl Diphosphate Isomerase. *Journal Of Bacteriology*, 181:4499-4504.
5. Jachim Hans, Bettina Hause, Dieter Strack, and Michael H. Walter. (2004). Cloning, Characterization, and Immunolocalization of a Mycorrhiza-Inducible

- 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase in Arbuscule-Containing Cell of Maize. *Plant Physiology*, 134: 614-624.
6. **Michael Campbell, Frederick M. Hahn, C. Dale Poulter and Thomas Leastek.** (1998). Analysis of the isopentenyl diphosphate isomerase gene family from *Arabidopsis thaliana*. *Plant Molecular Biology* .36: 323-328.
  7. **Oliver Laule, Andreas Fürholz, Hur-Song Chang, Tong Zhu, Xun Wang, Peter B. Heifetz, Wilhelm Grisse, and Markus Lange.** (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci* .100 (11): 6866 - 6871.
  8. **P.D. Matthews, E. T. Wurtzel.** (2000). Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl Microbiol Biothechno.* 153: 396-400.65
  9. **Zairen Sun, Francis X. Cunningham, Jr., and Elisabeth Gantt.** (1998). Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte . *Proc Natl Acad Sci* .95 (19): 11482 - 114881 .
  10. **Jonathan E. P., Toni M. K., Gerd Hause, Maja Raschke, Wenyun Gao, and Meinhard H. Z.** (2004).