

中国农业科学院

Chinese Academy of Agricultural Science

紫花苜蓿油菜素内酯基因*DWF4*的生物
信息学分析

Bioinformatics analysis of
brassinolide *DWF4* in *Medicago Sativa*

汇报人：刘 文 文

日 期：20180623



主要内容



目的及背景



基因功能分析



蛋白结构分析



总结

研究背景——紫花苜蓿

紫花苜蓿被誉为“牧草之王”，是世界上分布最广、最古老的栽培牧草，也是我国种植面积最大的人工牧草。而产量是紫花苜蓿长期以来的主要育种目标之一。



研究背景——DWF4基因的发现

1998年，Choe等人就从一个拟南芥T-DNA插入突变体中鉴定得到DWF4基因，DWF4基因的突变体具有许多特别明显的表型，首先突变体植株生长速度相对较缓，它的叶片小而聚集。（Azpiroz et al.1998）



研究背景——DWF4基因功能

DWF4基因在模式植物拟南芥当中是编码细胞色素氧化酶**CYP90B 1**，是**BR**（油菜素内酯）合成限速酶基因。

Campestanol和6-deoxo cathasterone能够使突变体表型恢复，其他中间产物和植物激素都不可以，该试验说明DWF4基因所编码的细胞色素氧化酶在BR合成路径中催化的是C-22羟化反应(Choe et al 1998)。

在拟南芥DWF4突变体中定量检测到有较高含量的C-22羟化反应的前体物质的累积，表明CYP90B 1主要是用来催化C-22羟基化反应过程的。

DWF4基因通过异位表达能够显著提高植物体内羟基化BR（油菜素内酯）的含量。

研究目的及意义

BR作为一种发现较晚的一种植物激素，对其在植物体内的功能的重要性是目前公认的，但其一些功能的具体作用机制还不甚清楚。

植物**BR**的研究主要集中于产量、品质以及植物性状的改良等。由于对**BR**的合成过程的研究已经较为明朗，目前对**BR**的研究主要集中在**BR**生物合成的调控和信号转导上，且研究对象多是集中在模式植物拟南芥、水稻等，而**对于紫花苜蓿的BR合成及信号转导方面的研究较少。**

紫花苜蓿具有高产、优质、抗逆性强、蛋白质含量高和适口性好等特点，因此，对紫花苜蓿进行**BR**的分子生物学研究意义十分重大。

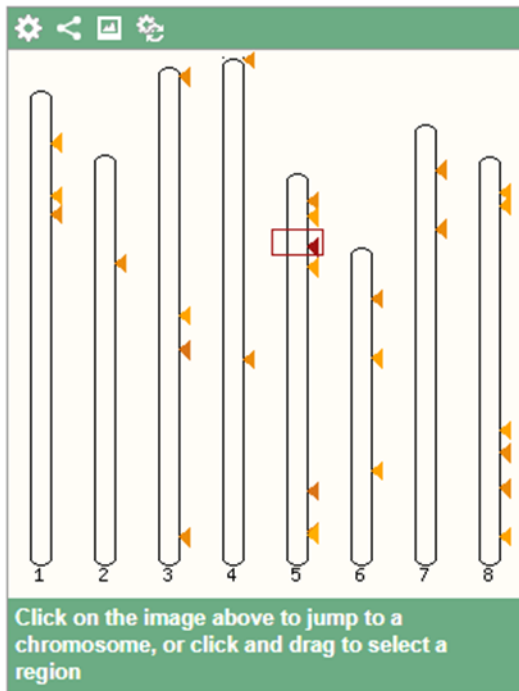
基因分析

DWF4基因的基本信息：紫花苜蓿中没有该基因登录号

物种	GeneID	拉丁名	登录号	长度
拟南芥	824229	<i>Arabidopsis thaliana</i>	NM_001339479.1	2092
玉米	100191562	<i>Zea mays</i>	XM_008669815.3	2052
胡杨	105110409	<i>Populus euphratica</i>	NM_001304313.1	1672
陆地棉	107896149	<i>Gossypium hirsutum</i>	NM_001326836.1	1565
紫花苜蓿		<i>Medicago sativa</i>		2302

基因分析

DWF4基因的基本信息（因紫花苜蓿基因组尚未发表，以蒺藜苜蓿为例）
以紫花苜蓿DWF4蛋白序列在蒺藜苜蓿基因组中进行BlastP比对（Ensemble数据库中进行）

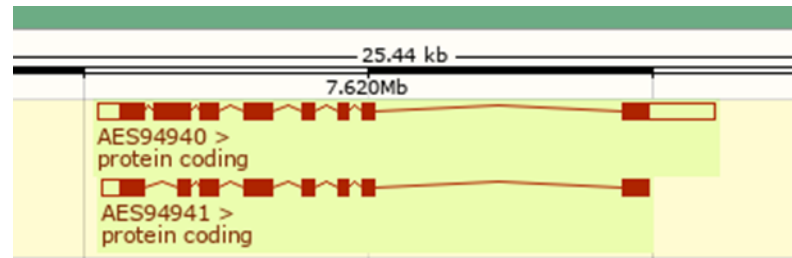


Description

cytochrome P450 family protein

Location

[Chromosome 5: 7,617,615-7,623,052](#) forward strand.

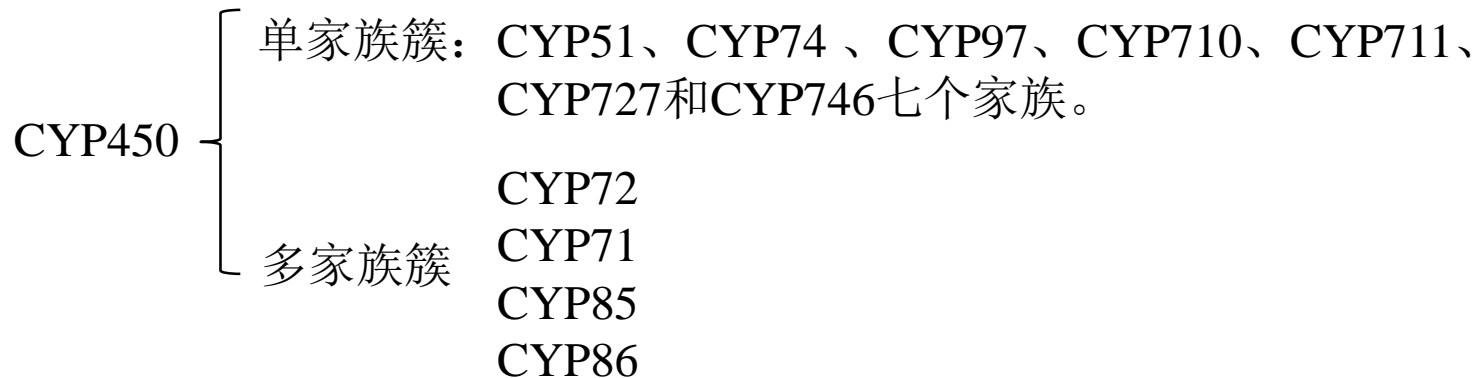


基因分析

将紫花苜蓿DWF4蛋白序列在Uniprot数据库中进行BlastP比对，结果如下：

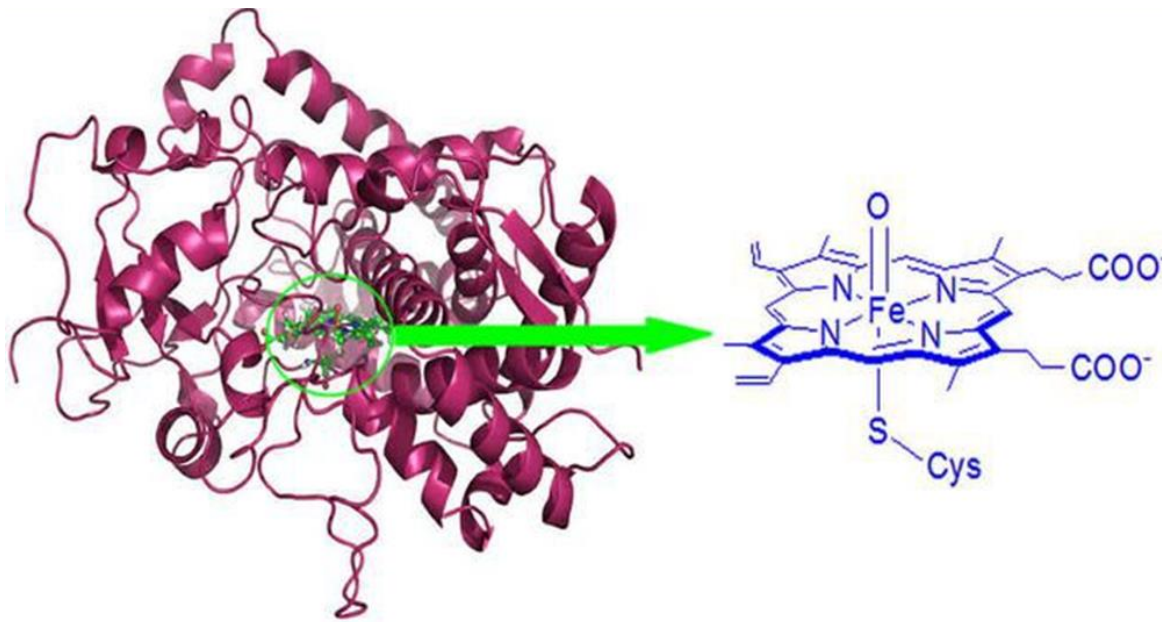
登录号	蛋白名称	物种	拉丁名	Identity
G7KDI9	Cytochrome P450 family protein	蒺藜苜蓿	<i>Medicago truncatula</i>	96.5%
A0A1S2YYQ0	cytochrome P450 90B1-like	鹰嘴豆	<i>Cicer arietinum</i>	90.7%
A4PBE7	Cytochrome P450 enzyme	豌豆	<i>Pisum sativum</i>	89.3%
A0A151TI98	Cytochrome P450 90B1	木豆	<i>Cajanus cajan</i>	85.1%

细胞色素P450(CYP450)是一类超基因家族编码的血红素单加氧酶，参与生物体内多种催化反应，主要分为生物合成和生物解毒两大类功能。



基因分析

CYP450存在保守的血红素结合域(Heme-binding region)FxxGxRxCxG, 同时也是判断一个基因是否属于CYP450家族的依据。



基因分析

同时以DWF4为基因名搜索，搜索其他物种中相关同源基因，结果如下：

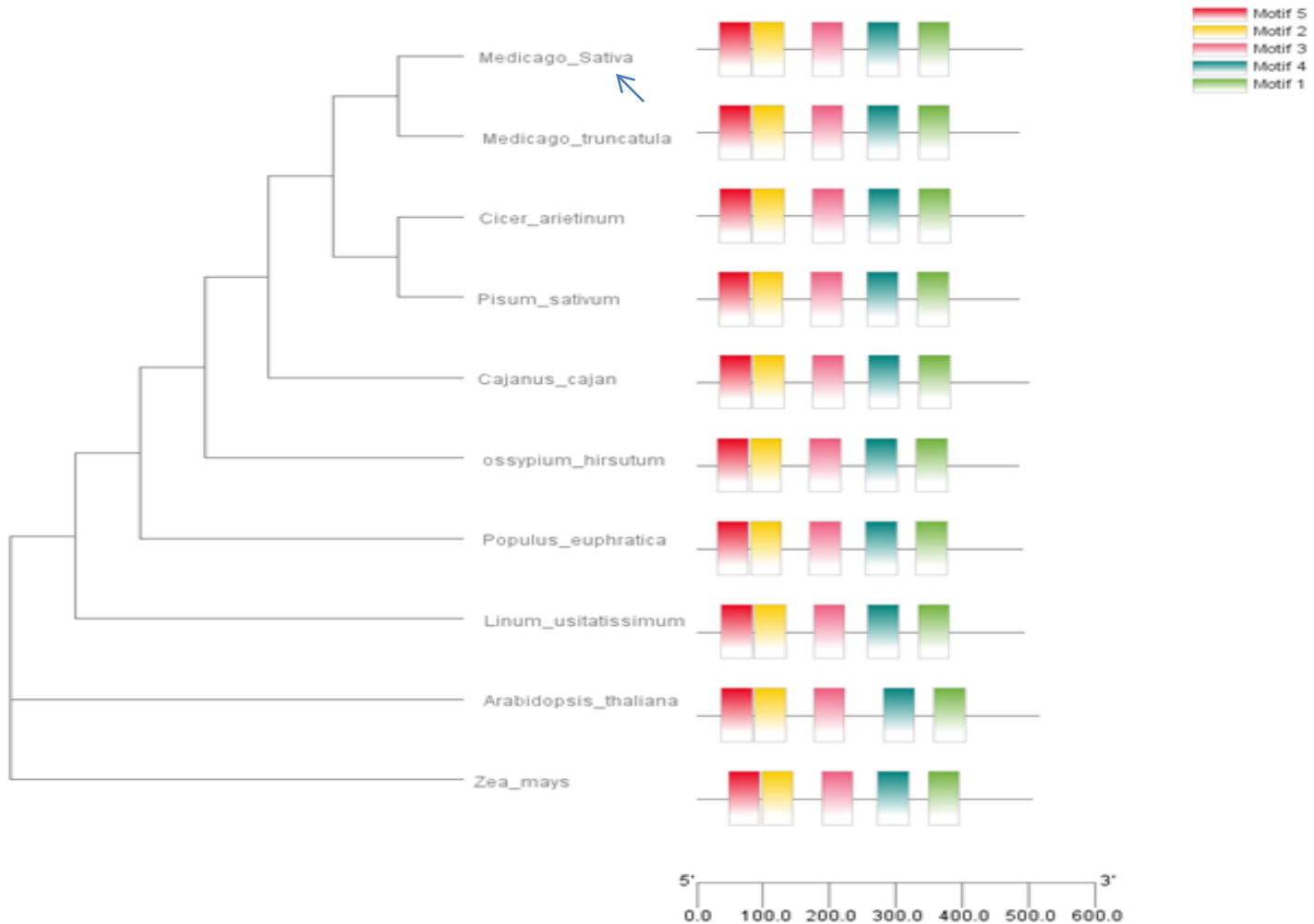
登录号	蛋白名称	物种	拉丁名
O64989	Cytochrome P450 90B1	拟南芥	<i>Arabidopsis thaliana</i>
D7EYE5	Steroid 22-alpha-hydroxylase protein	玉米	<i>Zea mays</i>
G8D3C4	Steroid 22-alpha hydroxylase	胡杨	<i>Populus euphratica</i>
A0A1W4	Steroid 22-alpha-hydroxylase	陆地棉	<i>Gossypium hirsutum</i>
A0A2D1GVF1	Cytochrome P450 monooxygenase	亚麻	<i>Linum usitatissimum</i>

将以上蛋白序列下载并进行多序列比对和进化树分析，在MEME网站（<http://meme-suite.org>）中对其进行保守结构域分析，发现其有5个比较保守的结构域，motif如下



基因分析

进化树及motif综合结果如下：



- 一级结构分析
- 二级结构分析
- 亚细胞定位
- 三级结构分析

蛋白结构分析——一级结构分析

紫花苜蓿中的DWF4蛋白分子式为 C2544H3946N674O695S20。

总共有 489 个氨基酸，相对分子质量为 55.73kDa，理论等电点为 9.0，带负电的残基数(Asp + Glu) 为 48，带正电荷的残基数(Arg + Lys) 为 57。

DWF4蛋白出现频率较高的氨基酸残基为 Leu(11.7%)、Lys(7.4%)、Ile(7.0%)，且不具有 Sec 与Pyl。

不稳定系数 40.87，可以推测此蛋白质不稳定。

<https://web.expasy.org/protparam/>

Number of amino acids: 489

Molecular weight: 55734.61

Theoretical pI: 9.00

Amino acid composition:

CSV format

Ala (A)	30	6.1%
Arg (R)	21	4.3%
Asn (N)	28	5.7%
Asp (D)	16	3.3%
Cys (C)	7	1.4%
Gln (Q)	14	2.9%
Glu (E)	32	6.5%
Gly (G)	33	6.7%
His (H)	17	3.5%
Ile (I)	34	7.0%
Leu (L)	57	11.7%
Lys (K)	36	7.4%
Met (M)	13	2.7%
Phe (F)	30	6.1%
Pro (P)	22	4.5%
Ser (S)	32	6.5%
Thr (T)	21	4.3%
Trp (W)	10	2.0%
Tyr (Y)	14	2.9%
Val (V)	22	4.5%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 48

Total number of positively charged residues (Arg + Lys): 57

蛋白质结构分析——二级结构预测

```

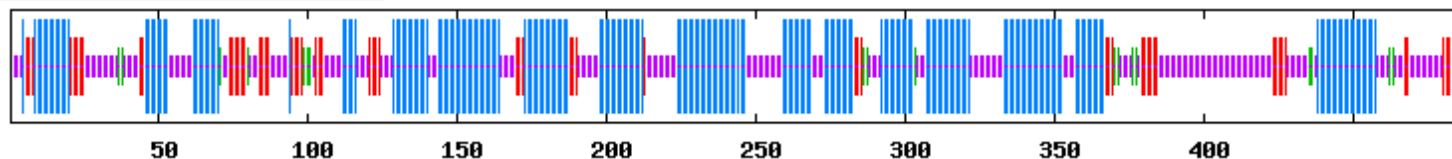
      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
MSDSDITFYFLSSIFSLLIFIFFLIKTGHAKPKLNLPPGRMGWPFIGETIGYLKPYSAITIGKFMEQHIA
cccccheeehhhhhhhhhhheeeeecccccccccccttccccceehhhhhhhhhccccccccchhhhhhh
RYGKIYKSNLFGGPTIVSADAGLNRFILQNEGKLFECSSYPSSIGGILGKWSMLVLVGMHRDMRNLISLNF
htceeeeeetcccccccccccccheeeettccccccccchhhhhccccceeeccccchhhhhhhhhhh
LCHARLRTHLLKEVEKHSRLVSSWKEKTTFAAQDEAKFTFNMAEHIMSLQPGKIETENLKKEYVTFM
hccccchhhhhhhhhhhhhhhhhhhhhccccceehhhhhhhhhhhhhheeeccccccccchhhhhhhhhhh
KGVVSAPLNFPGTAYWRALKSRCTILKFIKMEERIKRMQEGNENSEENDLLNwLKHNSNLSTEQILD
hhheccccccccchhhhhhhhhhhhhhhhhhhhhhhccccccccccccchhhhhhhhhhhccccchhhhhhh
ILSLLFAGHETSSVSIALAIYFLPGCPQAILQLREEHKEIARAKKLAGE TELTWEDYKKMEFTHCVVNET
hhheeeettccccchhhhhhhhhhhhtcccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
LRLGNVVRFLHRKALKDVRVYKGYDIPCGWVKVLPVIAAVHLDPLLLFDQPQHFNPWRWQNNNGNNGPNSASG

```

SOPMA :

Alpha helix	(Hh)	:	218 is	44.58%
3_{10} helix	(Gg)	:	0 is	0.00%
Pi helix	(Ii)	:	0 is	0.00%
Beta bridge	(Bb)	:	0 is	0.00%
Extended strand	(Ee)	:	61 is	12.47%
Beta turn	(Tt)	:	18 is	3.68%
Bend region	(Ss)	:	0 is	0.00%
Random coil	(Cc)	:	192 is	39.26%
Ambiguous states (?)		:	0 is	0.00%
Other states		:	0 is	0.00%

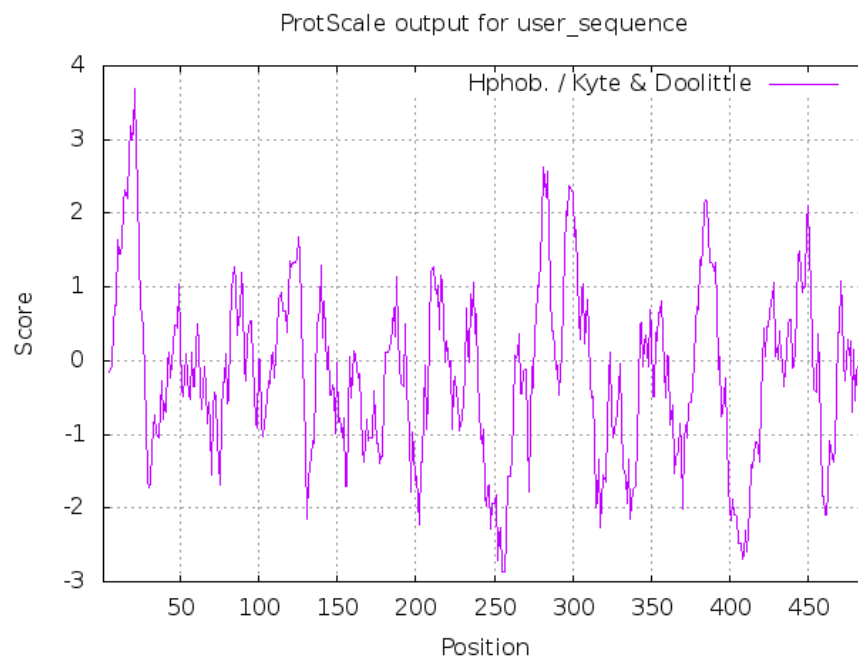
α 螺旋
 3_{10} 螺旋
 Pi螺旋
 B桥
 延伸链
 B转角
 弯曲区域
 无规则区域
 不明确的状态
 其他状态



蛋白结构分析——二级结构预测

亲疏水性预测分析：氨基酸表现为亲水性时数值为负值，表现为疏水性时数值为正，数值绝对值越高表明亲疏水性越强。

该蛋白总平均亲水性为：-0.160



蛋白结构分析

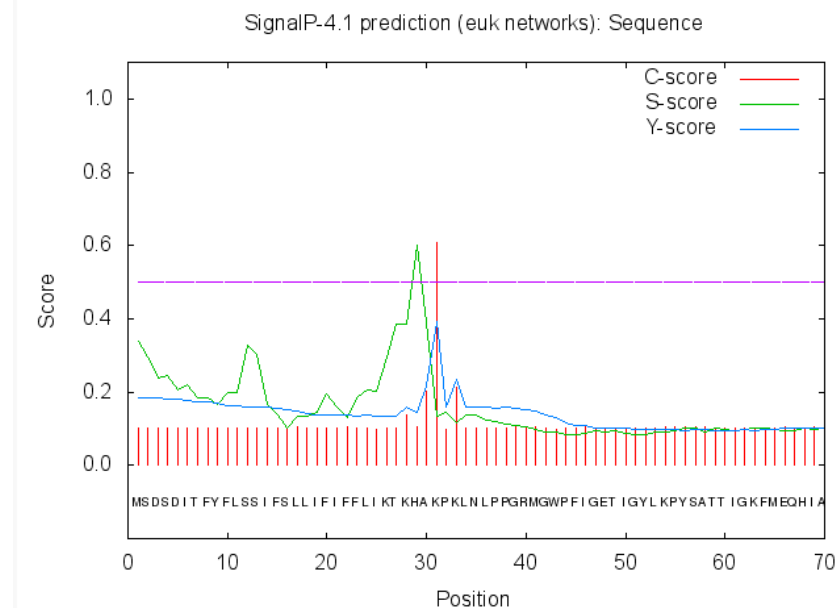
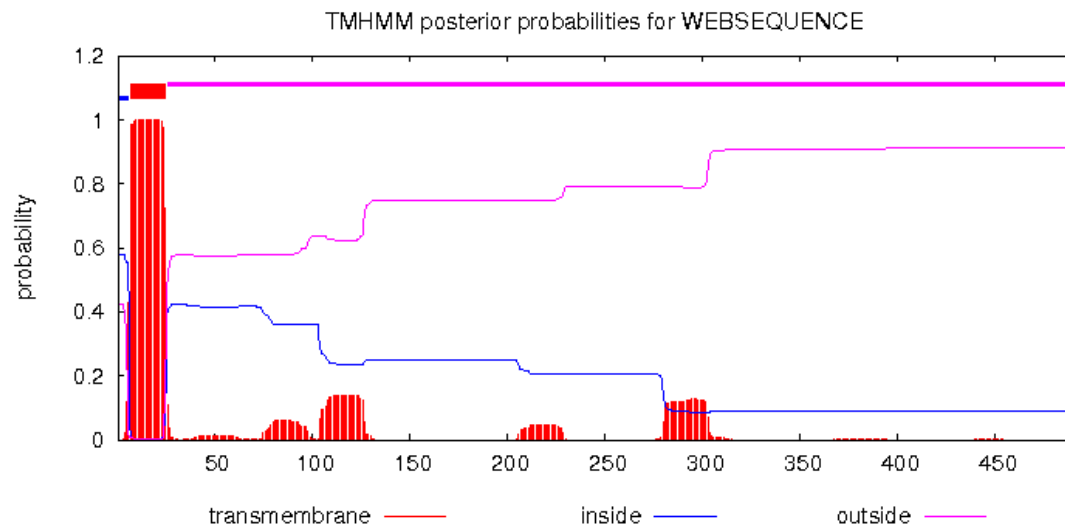
蛋白的跨膜结构分析：膜外蛋白，没有较为明显的跨膜结构。

红色柱子表示可能是跨膜区域的可信度，综合来说，0到50为跨膜蛋白，50到150可能性各一半，所以并不确定。150以后部分可能是蛋白质的膜外部分。

TMHMM Server 2.0

<http://www.cbs.dtu.dk/services/TMHMM/>

运用SignalP — 4.1预测信号肽



蛋白结构分析——亚细胞定位

进行BLAST序列比对，在找到同源相似较近的物种进行亚细胞定位。发现*Medicago truncatula*（蒺藜苜蓿）的细胞色素P450 90B1，与目的蛋白相似度较高，占97%左右。



Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	cytochrome P450 90B1 [Medicago truncatula]	977	977	100%	0.0	97%	XP_003611982.2
<input type="checkbox"/>	PREDICTED: cytochrome P450 90B1-like [Cicer arietinum]	912	912	100%	0.0	91%	XP_004512038.1
<input type="checkbox"/>	cytochrome P450 90B1 [Cajanus cajan]	864	864	100%	0.0	85%	XP_020218337.1
<input type="checkbox"/>	cytochrome P450 enzyme [Pisum sativum]	858	858	99%	0.0	88%	BAF56239.1

From June 20, 2018 all traffic will be automatically redirected to HTTPS. More information or view this page using https

UniProtKB - O64989 (C90B1_ARATH)

Display

Entry

Publications

Feature viewer

Feature table

All None

Function

Names & Taxonomy

[BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)

[Feedback](#) [Help video](#)

Protein | **Cytochrome P450 90B1**

Gene | **CYP90B1**

Organism | *Arabidopsis thaliana* (Mouse-ear cress)

Status | Reviewed - Annotation score: ●●●●●● - Experimental evidence at protein level¹

Function¹

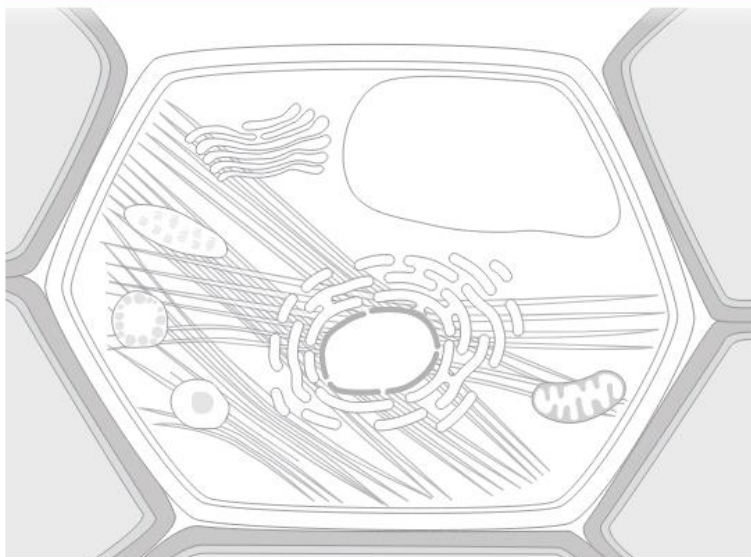
Catalyzes the C22-alpha-hydroxylation step in brassinosteroids biosynthesis. Converts campestanol to 6-deoxocathasterone and 6-oxoca

蛋白结构分析

在uniprot中进行检索，发现了该蛋白的注释信息
亚细胞定位：在膜上，主要是内质网。

TAIRⁱ locus:2101704 AT3G50660

Subcellular locationⁱ



Graphics by Christian Stolte. Source: COMPARTMENTS

UniProt annotation

GO - Cellular component

Other locations

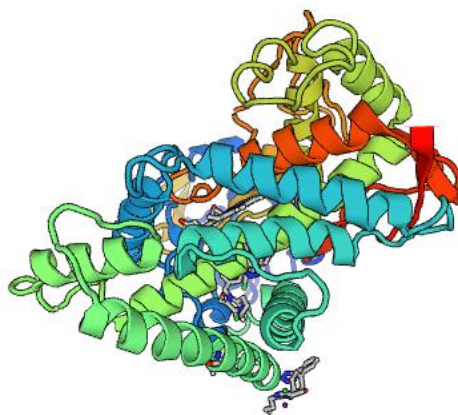
Membrane ⁱ Curated ; Single-pass membrane protein ⁱ
Curated

蛋白质结构分析——三维结构分析

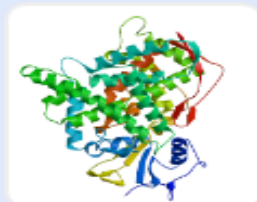
Cytochrome P450 2C9

Seqres	MAKKTSGRGKLP	PGPTPLPVIGNILQIGIKDISKSLTNLSKVYGPVFTLYFGLKPIVVLHGYEAVKEALIDLGEFEFSGRG	80			
5x23.1.A	-----KLPPGPTPL	PGPTPLPVIGNILQIGIKDISKSLTNLSKVYGPVFTLYFGLKPIVVLHGYEAVKEALIDLGEFEFSGRG	80			
Seqres	IFPLAERANRGFGIVFSNGKKWKEIRRFSLM	TLRNFNGMGRSIEDRVQEEARCLVEELRKTASPCDPTFILGCAPCNVI	160			
5x23.1.A	IFPLAERANRGFGIVFSNGKKWKEIRRFSLM	TLRNFNGMGRSIEDRVQEEARCLVEELRKTASPCDPTFILGCAPCNVI	160			
Seqres	CSIIIFHKRFDYKDQQFLNLMKLNENIKILSS	PWIQICNNSPIIDYFPGTHNKLLKNVAFMKSYLEKVKHQESMDMN	240			
5x23.1.A	CSIIIFHKRFDYKDQQFLNLMKLNENIKILSS	PWIQICNNSPIIDYFPGTHNKLLKNVAFMKSYLEKVKHQESMDMN	240			
Seqres	NPQDFIDCFLMKMEKEKHNQPSFTIESLENT	AVDLFGAGTETTSTTLRYALLLLKHPEVTAKVQEEIERVIGNRSPC	320			
5x23.1.A	NPQDFIDCFLMKMEKEKHNQPSFTIESLENT	AVDLFGAGTETTSTTLRYALLLLKHPEVTAKVQEEIERVIGNRSPC	320			
Seqres	MQDRSHMPYTD	AVVHEVQRYIDLLP	TS	SLPHAVTCDIKFRNYLIPKGT	TILISLTSVLDHNKEFPNPEMFDPHHFLDEGGN	400
5x23.1.A	MQDRSHMPYTD	AVVHEVQRYIDLLP	TS	SLPHAVTCDIKFRNYLIPKGT	TILISLTSVLDHNKEFPNPEMFDPHHFLDEGGN	400
Seqres	FKKSKYFMPFS	SAGKRICVGEALAGMELFLFLTSILQNFNLKSLVDPKNLDTTPV	VNGFTSVPPFYQLCFPIHHHH	476		
5x23.1.A	FKKSKYFMPFS	SAGKRICVGEALAGMELFLFLTSILQNFNLKSLVDPKNLDTTPV	VNGFTSVPPFYQLCFPIHHHH	472		

细胞色素P450 2C9结构相近



蛋白质结构分析——三维结构评价



Model 01 ▾

Structure
Assessment

Oligo-State ⓘ

Monomer

Ligands ⓘ

None

GMQE ⓘ

0.59

QMEAN ⓘ

-4.68 ↕

Global Quality Estimate

QMEAN -4.68

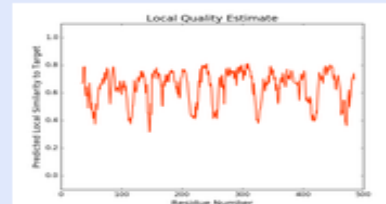
C β -2.86

All Atom -1.60

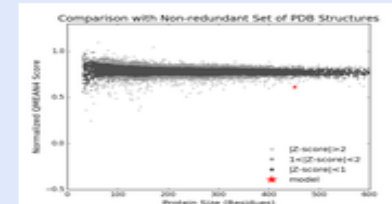
Solvation -0.03

Torsion -4.24

Local Quality Estimate



Comparison



Template Seq Identity Coverage

5x23.1.A 21.38%

Description

Cytochrome P450 2C9

Model-Template Alignment

Model_01	MSDSDITFYFLSSIFSLIFIFFLIKTKHAKPKLNLPPGRMGWPFIFGETIGYLPKPY\$ATTIGKFMEOHIARYGKIYKSNL	80
5x23.1.A	-----TSGRGKLPPTPEIPVIGNQLQIG----IKIISKSLTNLSKVYGEVFTLY	51
Model_01	FGGPTIV\$ADAGLNRFILQNEGKLFECSTP--SSIGGILGKWSMLVLVGDMDHRDMRNISLNFLCHA--RLRTHLLKEVEKH	157
5x23.1.A	GIKPIVVVHGVEAVKEALIDLGEEFSSGSIFFLAERANRGGFIVFSNGKKWKEIRRFSLMTLRNFGMGKRS--TEDRVQEE	130
Model_01	SRLVVSSWKE--KTTEAAQDEAKKFTFNLMAEHIMSLQPG--KIETENLKKEYVTFMKGVV-----SAP-L--NPPG	222
5x23.1.A	ARCLVEELRKTKASFDPTFILGCAPCNVICSIIFHKRFDYKDAQFLNLMKELNENIKIDSSPWIQICNDFSPIIDYFPG	210
Model_01	TAYWRALKSRCTILKFIIEGKMEERIKRMQEGNENSEENDLLNWVLKH-----S-NLSTEQILDILSLLFAGHETSS	293
5x23.1.A	T-HNKLKLNVAFMKSYILEKVKHEHQESMDM----NNPQDEFIDCFLMKMEKPKHNQPSEFTIESLENTAVDLFGAGTETTS	285
Model_01	VSIALAIYFLPGCPQAILQLREEHKEIARAKKLAGETELTWEDYKMEFTHCVVNETLRLGNVVR-FLHRRKALKDVRYPKG	372
5x23.1.A	FTLRVALLLLRHPEVTAQVQEEIERVDS----RNRS-PCMQDRSIMEYTDVVVHEVQRYDILLPTSLPHTVTCDIKRN	360
Model_01	YDIPCGWKVLPVIAAVHLDPLLFDQPQHFNPRWQNNGNNGPNSSASGNSNIFLPPFGGGPRLCAGSELAKLEMAVFIHHL	452
5x23.1.A	YLDIPKGTITLDSLTSVLDHNKEFPNPEMFDPHHFDDEGGN-----KSKYFMPFSA GKRICVGEALAGMELFLFLTSTI	434
Model_01	ILNYNWELTDNNDQ-AF--AY-PFVDFPKGLQIRVQAHSII	489
5x23.1.A	LQNFENLKEVLDVDPKNLDTTFWVNGFDSPPFFYQLCFLE----	471



致谢

衷心感谢罗老师一学期以来严格、无私、高质量的教导！

祝愿罗老师身体健康，工作愉快！

感谢所有组员对学习、讨论等各项工作的配合！

感谢各位同学的聆听！

希望所有同学在以后的生活中一切顺利！

请老师、同学批评指正！
谢谢！



中国农业科学院
China Academy of Agricultural Science