

**LOGO**

# 西瓜a基因克隆与序列分析

**13组** 任毅 吕红豪 王燕 张玉苗





# 对植物性别研究是永恒的科学命题

葫芦科是植物性别研究的模式生物





## 葫芦科（西瓜）七种性型及基因型

分类	性型	基因型	符号	西瓜材料
单一性型	全雌系 <b>Gynoecious</b>	<b>A-gygy</b>	♀	新红宝雌性突 变体 ( <b>XHBFGM</b> )
	全雄系 <b>Androecious</b>	?	♂	
	完全花系 <b>Hermaphrodite</b>	<b>aagygy</b>	♀	
多态性型	雌雄同株异花系 <b>Monoecious</b>	<b>A-GY-</b>	♀ ♂	新红宝母本 ( <b>XHBF</b> )
	完全花雌花系 <b>Gynomonoeocious</b>		♀ ♀	
	完全花雄花系 <b>Andromonoecious</b>	<b>aaGY-</b>	♂ ♀	苏联3号，阿柯 克孜外
	三性花系 <b>Trimonoecious</b>		♀ ♂ ♀	



转录因子参与调控gg基因

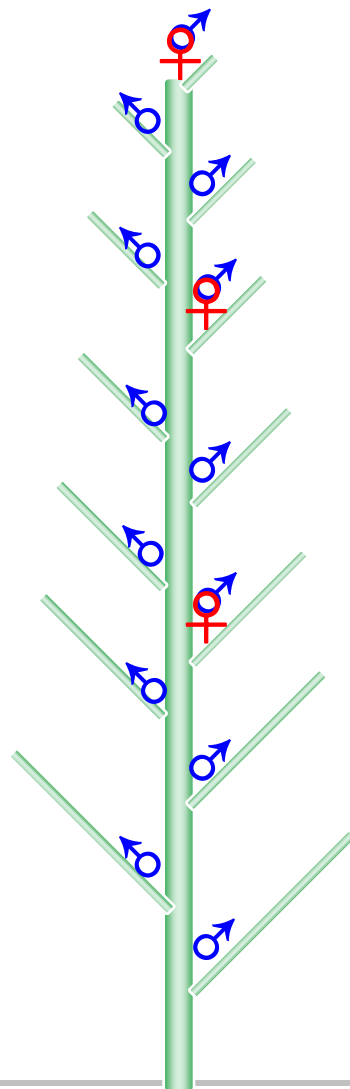
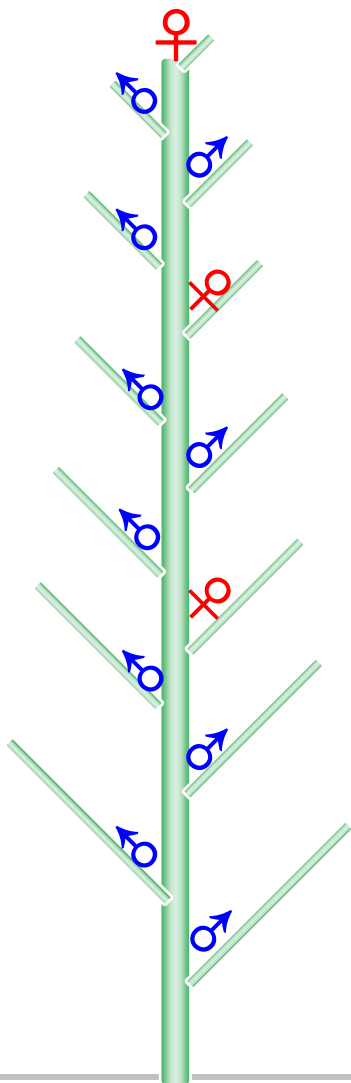


## 西瓜性型及果实





## 本研究材料性型模式图

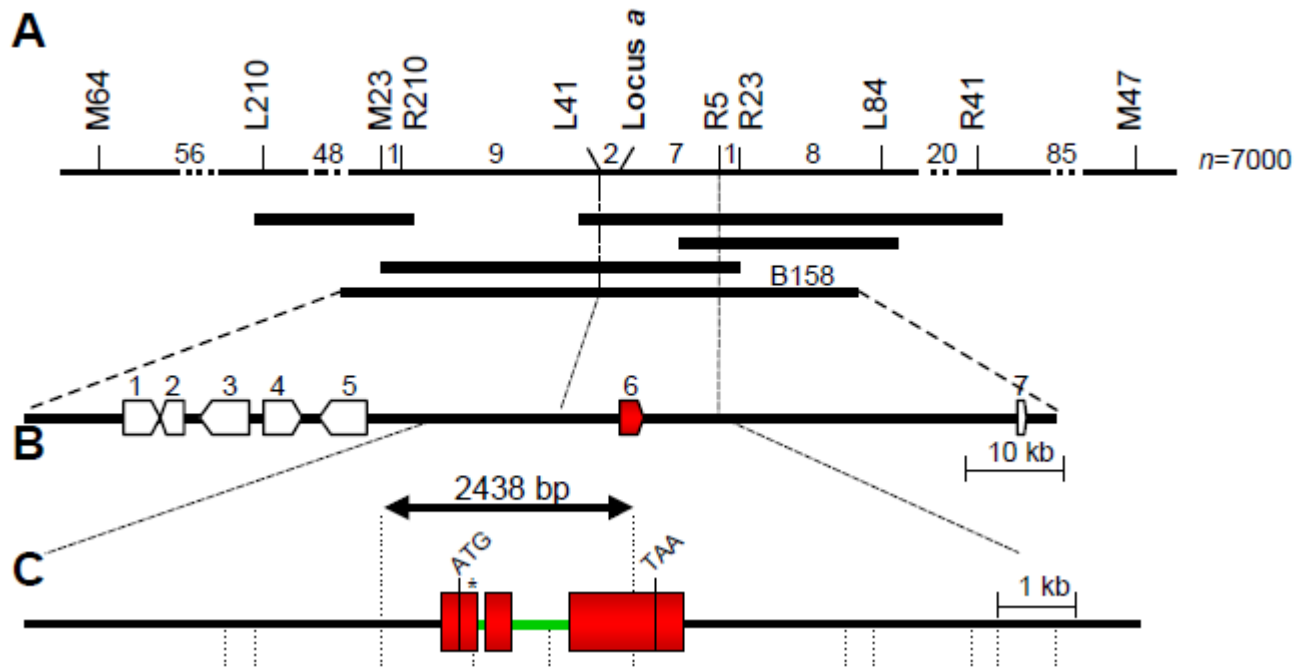


雌雄同株异花系 **Monoecious**

完全花雄花系 **Andromonoecious**



# 基因克隆策略







# DNA外显子、内含子预测及蛋白翻译: GENESCAN、FGENESH

## Gene Finding: Gene models construction, Splice sites, Protein coding exons

Load a local file with sequence in Fasta format. Then select the program and perform search.

Local file name:  Melon\_A-a\_gene\_DNA.fasta

Alternatively: [paste nucleotide sequence](#).

### Choose search method, closest organism and click Perform search:

- 
- FGENESB** / Prediction coding genes in bacterial metagenomic sequences (both chains) [[Help](#)]
  - PROMB** / Prediction of bacterial promoters [[Help](#)]
  - Organism:  Human  Mouse  Drosophila  C.elegans  Yeast (S.c.)  Dicots (Arabidopsis)
    - Monocots (Corn, Rice, Wheat, Barley)  Schizosaccharomyces pombe  Neurospora crassa
  - FGENESH** / HMM based (select any organism from above list)
  - Gene structure prediction (multiple genes, both chains) [[H](#)] (Superfast variant provided for human only: 50-100 times faster)
  - BESTORF** - Finding potential coding fragment EST/mRNA (Human/Drosophila/Plant (Dct)) [[H](#)]
  - TSSP\_TCM** - Prediction of Promoters in Plant sequences [[H](#)]

Predicted protein(s):

```
>FGENESH: 1 3 exon (s) 3651 - 5659 445 aa, chain +
MAIEIDIEQNPTVELSRIGTSETHGEDSPYFAGWKAYDEDPYNSTNPSGVIQMGLVENQ
VSFDLLEEYLEENCEGEGNYLNSGFRENALFQDYHGLFSFRSAMGSMFEEIRGGRAKFD
NRVVLTAGATAANELLTFILANPGDALLVPTPYYPGFDRDLRWRTGVKIVPIHCDSSN
FITPKALEEAYNSAMEMKIKVRGVLITNPSNPLGATIQRSTIEDILDFVTRKNIHLVSD
IYSGSVFSSAEFTSVAEVLERSRSYKNAERVHIVYSLSKDLGLPGFRIGTIYSYNDKVV
ARRMSSFTLISSQTQRFLASMLSNRKFTEKYIKMNRDRLKKRYEMIIEGLRTAGIECLE
NAGLFCWMNLSPLLKDKKTKEGEIEIWKRILKEVKLNISPGSSCHCSEPGWFRVCFAN
EKTLHVALDRIRRFMERMKKENEAN
```



## AA基因和aa基因由第364位一个碱基突变引起： C半胱氨酸→W色氨酸

Score = 923 bits (2385), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 443/444 (99%), Positives = 443/444 (99%), Gaps = 0/444 (0%)

```
Query 1 MAIEIEIEQNPTVELSQIGTSETHGEDSPYFAGWKAYDEDPYNETTNP SGVIQMGLAENQ 60
MAIEIEIEQNPTVELSQIGTSETHGEDSPYFAGWKAYDEDPYNETTNP SGVIQMGLAENQ
Sbjct 1 MAIEIEIEQNPTVELSQIGTSETHGEDSPYFAGWKAYDEDPYNETTNP SGVIQMGLAENQ 60

Query 61 VSFDLLEEYLEQNCEAEANCSGFRENALFQDYHGLLSFRTAMAGFMEEIRGGRAKFDPNR 120
VSFDLLEEYLEQNCEAEANCSGFRENALFQDYHGLLSFRTAMAGFMEEIRGGRAKFDPNR
Sbjct 61 VSFDLLEEYLEQNCEAEANCSGFRENALFQDYHGLLSFRTAMAGFMEEIRGGRAKFDPNR 120

Query 121 VVLTAGATAANELLT FILANPGDALLVPTPYYPGFDRDLRWRTGVKIVPIHCDSSNNFQI 180
VVLTAGATAANELLT FILANPGDALLVPTPYYPGFDRDLRWRTGVKIVPIHCDSSNNFQI
Sbjct 121 VVLTAGATAANELLT FILANPGDALLVPTPYYPGFDRDLRWRTGVKIVPIHCDSSNNFQI 180

Query 181 TPKALEEAYNTAMAMKIKVRGVLITNPSNPLGATIQRSTIEEILDFVTRKNIHLSVDEIY 240
TPKALEEAYNTAMAMKIKVRGVLITNPSNPLGATIQRSTIEEILDFVTRKNIHLSVDEIY
Sbjct 181 TPKALEEAYNTAMAMKIKVRGVLITNPSNPLGATIQRSTIEEILDFVTRKNIHLSVDEIY 240

Query 241 SGSVFSSAEFTSVAEVLESRGYKNAERVHIVYSLSKDLGLPGFRIGTIYSYNDKVVTTAR 300
SGSVFSSAEFTSVAEVLESRGYKNAERVHIVYSLSKDLGLPGFRIGTIYSYNDKVVTTAR
Sbjct 241 SGSVFSSAEFTSVAEVLESRGYKNAERVHIVYSLSKDLGLPGFRIGTIYSYNDKVVTTAR 300

Query 301 RMSSFTLISSQTQRFLASMLSNRKFTEKYIKMNRDLRKRKRYEMIEGLRTAGIQCLEGNA 360
RMSSFTLISSQTQRFLASMLSNRKFTEKYIKMNRDLRKRKRYEMIEGLRTAGIQCLEGNA
Sbjct 301 RMSSFTLISSQTQRFLASMLSNRKFTEKYIKMNRDLRKRKRYEMIEGLRTAGIQCLEGNA 360

Query 361 GLEFCWMNLSPLLKNKKL TREGEIEIWKRILKEVKNLISPGSSCHCSEPGWFRVCFANMSE 420
GLEFCWMNLSPLLKNKKL TREGEIEIWKRILKEVKNLISPGSSCHCSEPGWFRVCFANMSE
Sbjct 361 GLEFCWMNLSPLLKNKKL TREGEIEIWKRILKEVKNLISPGSSCHCSEPGWFRVCFANMSE 420

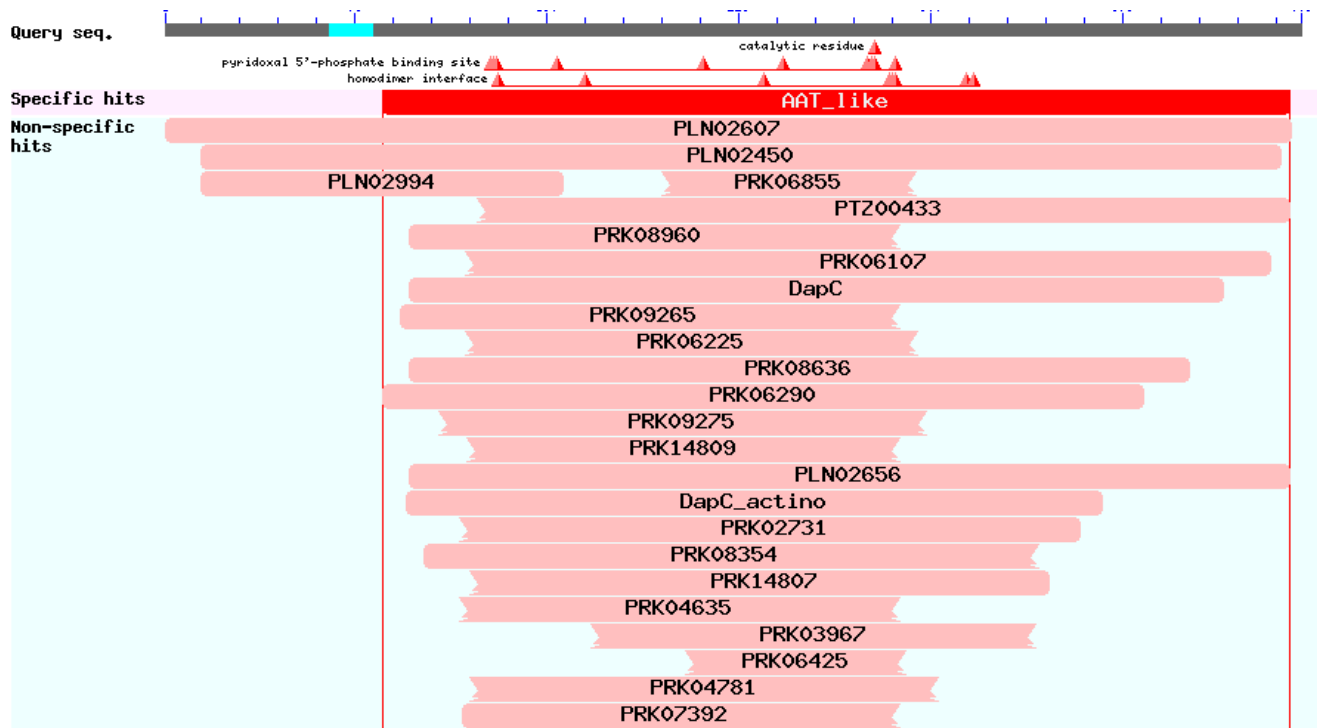
Query 421 KTLHVALERIHC FMERM RKEDEVN 444
KTLHVALERIHC FMERM RKEDEVN
Sbjct 421 KTLHVALERIHC FMERM RKEDEVN 444
```







## 蛋白注释：NCBI、InterPro Scan、PHAM

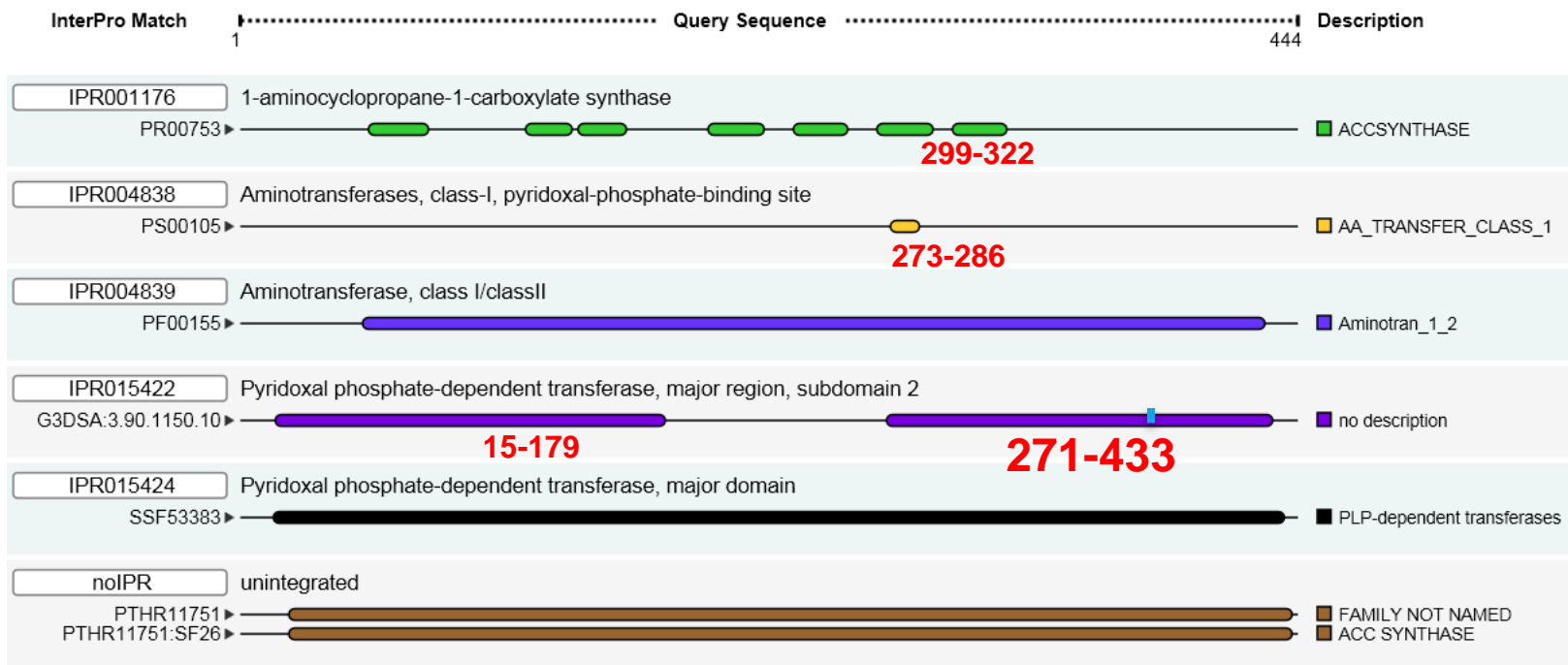


含有AAT\_like家族：天门冬氨酸氨基转移酶家族，是乙烯合成酶家族基因的特征。

同源基因中相似度为90%的均为乙烯合成酶基因，因此可进一步确定该基因为乙烯合成酶基因家族基因。



# InterPro Scan、PHAM注释结果: Pyridoxal phosphate( PLP, 磷酸吡哆醛) dependent aminotransferase superfamily



■ PRODOM ■ PRINTS ■ PIR ■ PFAM ■ SMART ■ TIGRFAMs ■ PROFILE  
■ HAMAP ■ PROSITE ■ SUPERFAMILY ■ SIGNALP ■ TMHMM ■ PANTHER ■ GENE3D

Family	Description
Aminotran_1_2	Aminotransferase class I and II
#HMM	dvinLgsneylgdsgkptlpevakaek...galaggtlneygpigdlpelealakflg...rsek1klkreaavvvgGagaliealifl1kl1npgdeilvpdpt
#MATCH	++++L++n+ ++d l+e ++++ e + +++++ +++++gl +++++a+a f+ r+ + k++++ +vv++ Ga+a+++e+l+f+l +npgd+ lvp+p+
#PP	689*****...777777776655558899*****.*****.*****.9*****
#SEQ	IQMGLEAENQVSFD----LLEEYLEQNCEaeaNCSGFRENALFQDYHGLLSFRTAMAGFMEeiRGGRAKFDPN-RVVLTAGATAANELLTFIL-ANPGDALLVPTPY



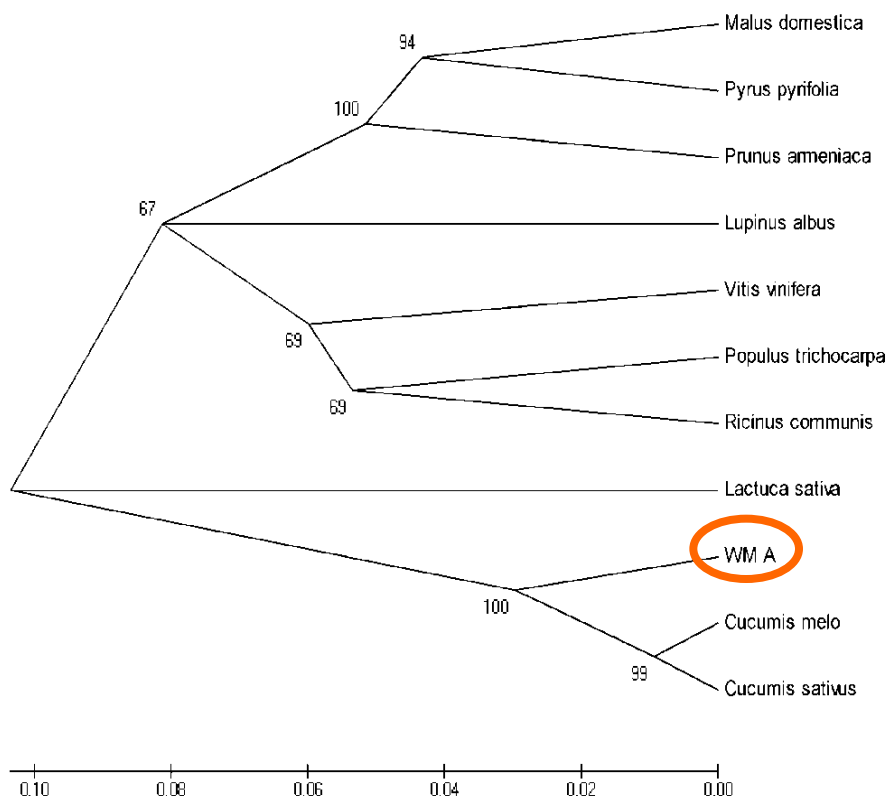
## 同源基因

Filter · Overview · Results · Job information Customize order											Page 1	of 10   Next
Alignments	Entry	Entry name	Status	Protein names	Organism	Length	Identity	Score	E-value	Gene names		
<input type="checkbox"/>	B5LHP3	B5LHP3_CUCME	★	1-aminocyclopropane-1-carboxylic acid synthas...	Cucumis melo (Muskmelon)	445	93.0%	2,163	0.0	ACS-7		
<input type="checkbox"/>	B5LHP4	B5LHP4_CUCME	★	1-aminocyclopropane-1-carboxylic acid synthas...	Cucumis melo (Muskmelon)	445	93.0%	2,159	0.0	ACS-7		
<input type="checkbox"/>	A8ASI7	A8ASI7_CUCSA	★	1-aminocyclopropane-1-carboxylate synthase	Cucumis sativus (Cucumber)	445	92.0%	2,142	0.0	ACS2		
<input type="checkbox"/>	C7BFM6	C7BFM6_CUCSA	★	1-aminocyclopropane-1-carboxylate synthase	Cucumis sativus (Cucumber)	445	92.0%	2,133	0.0	ACS2		
<input type="checkbox"/>	B9H8V4	B9H8V4_POPTR	★	1-aminocyclopropane-1-carboxylate	Populus trichocarpa (Western balsam poplar) (Populus balsamifera subsp. trichocarpa)	440	82.0%	1,877	0.0	ACS6 POPTRDRAFT_762657		
<input type="checkbox"/>	F6HGS0	F6HGS0_VITVI	★	Putative uncharacterized protein	Vitis vinifera (Grape)	445	82.0%	1,864	0.0	VIT_11s0016g02560		
<input type="checkbox"/>	A5AM70	A5AM70_VITVI	★	Putative uncharacterized protein	Vitis vinifera (Grape)	453	82.0%	1,864	0.0	VITISV_029694		
<input type="checkbox"/>	B9RKB7	B9RKB7_RICCO	★	Acc synthase, putative	Ricinus communis (Castor bean)	440	81.0%	1,851	0.0	RCOM_1048130		
<input type="checkbox"/>	Q9SEK0	Q9SEK0_LUPAL	★	1-aminocyclopropane-1-carboxylate synthase 4	Lupinus albus (White lupin) (Lupinus termis)	446	80.0%	1,849	0.0	ACS4		
<input type="checkbox"/>	B9N5G2	B9N5G2_POPTR	★	1-aminocyclopropane-1-carboxylate	Populus trichocarpa (Western balsam poplar) (Populus balsamifera subsp. trichocarpa)	446	80.0%	1,845	0.0	ACS5 POPTRDRAFT_582947		
<input type="checkbox"/>	Q9SEJ6	Q9SEJ6_LUPAL	★	1-aminocyclopropane-1-carboxylate synthase 5	Lupinus albus (White lupin) (Lupinus termis)	448	80.0%	1,816	0.0	ACS5		
<input type="checkbox"/>	Q84QH5	Q84QH5_LACSA	★	Putative ACC synthase 2	Lactuca sativa (Garden lettuce)	445	79.0%	1,814	0.0	ACS2		
<input type="checkbox"/>	Q1HAW0	Q1HAW0_MALDO	★	1-aminocyclopropane-1-carboxylic acid synthas...	Malus domestica (Apple) (Pyrus malus)	446	77.0%	1,812	0.0	MdACS3c		
<input type="checkbox"/>	F8RWY2	F8RWY2_PRUAR	★	1-aminocyclopropane-1-carboxylate synthase 3	Prunus armeniaca (Apricot)	440	78.0%	1,810	0.0	ACS3		
<input type="checkbox"/>	Q9SXN2	Q9SXN2_PYRPY	★	1-aminocyclopropane-1-carboxylate synthase	Pyrus pyrifolia (Chinese pear) (Pyrus serotina)	446	77.0%	1,802	0.0	pPPACS2		
<input type="checkbox"/>	Q6T2D9	Q6T2D9_PYRCO	★	1-aminocyclopropane-1-carboxylate synthase 3	Pyrus communis (Pear) (Pyrus domestica)	446	76.0%	1,800	0.0	ACS3a ACS3		





## MAGA5.0构建系统发育树





# 同源基因多序列比对

```

Apple      MGIEIEQQQQPSVGLSKIIVSNTHGEDSPYFAGWKAYDENPYDESSNPSGVIQMGLAENQ 60
Chinese    MAIDIEQQQQPSPGLSKIIVSDTHGEDSPYFAGWKAYDENPYHESSNPSGVIQMGLAENQ 60
Apricot    MAIEIE---QPSVGLSKIIVSDTHGEDSPYFAGWKAYDEDPYNQSTNPSGVIQMGLAENQ 57
Lupinus    MGIEIE---KCCVELSKIVANSETHGEDSPYFAGWKAYDEDPYHEHTNPSGVIQMGLAENQ 57
Western    MAIEIEQ---PSVGLSKVAVSETHGEDSPYFAGWKAYNEDPYDESSNPSGVIQMGLAENQ 57
Castor     MAIEIEQ---PAVGLSKIIVSETHGEDSPYFAGWKAYDEDPYHEYENPSGVIQMGLAENQ 57
Grape      MAIEIDQQ--PPVGLSRVAVSETHGEDSPYFAGWKAYDENPYDESKNPSGVIQMGLAENQ 58
Garden     MAIEIVQ---PSVGLSKIIVSETHGEDSPYFAGWKAYDENPYDEVHNP TGVIQMGLAENQ 57
Muskmelon  MAIEIDIEQNPTVELSRIGTSETHGEDSPYFAGWKAYDEDPYNSTNPSGVIQMGLAENQ 60
Cucumber   MAIEIEIEQNSSVELSRIGTSETHGEDSPYFAGWKAYDEDPYNSTNPSGVIQMGLAENQ 60
WM_A       MAIEIEIEQNPTVELSQIGTSETHGEDSPYFAGWKAYDEDPYNETTNP SGVIQMGLAENQ 60
          *.:*          *.:. *.:*****:.*.: *.:*****

Apple      VSFVLEKYLEEHSEASNWGAKASKGVSGFRENALFQDYHGLLSFRKAMASFMEQIRGGR 120
Chinese    VSFDLLEKXLEENSEASNWGSKGSKGVSGFRENALFQDYHGLLSFRKAMASFMEQIRGGR 120
Apricot    VSFDLLEKYLEEHSEASNWG---SKGVPCFRENALFQDYHGLVSRKAMASFMEQIRGGR 114
Lupinus    VSFDLLEKYLEEHSETVKWG---KRGRGFRENALFQDYHGLSFRKAMGSFMEQIREGR 113
Western    VSFDLLEEYLEQHSEASSWG---KSGPGFRENALFQDYHGLSFRQAMAGFMEQIRGGR 113
Castor     VSFDLLEEYLAKHSEASSWG---KGALSFRENALFQDYHGLSSFRQAMASFMEQIRGGR 113
Grape      VSFDLLEEFKQHSSEASSWG---NKISGFRENALFQDYHGLQSFRKAMASFMEQIRGGR 114
Garden     VSFDLLEEYLEANLEASNWG---QKVS GFRENALFQDYHGLQSFRKAMASFMEQVRGGK 113
Muskmelon  VSFDLLEEYLEENCEGEGNY----LNSGFRENALFQDYHGLSFRSAMGSFMEEIRGGR 115
Cucumber   VSFDLLEEYLEENCEGEGNY----LNSGFRENALFQDYHGLSFRSAMGSFMEEIRGGR 115
WM_A       VSFDLLEEYLEQNCAEAN-----CSGFRENALFQDYHGLLSFR TAMAGFMEEIRGGR 113
          ****:*.:*   : *          ***** ** **.* * : * *

Apple      AKFDPDRIVLTAGATAANELLT FIIAGPGDALLVPT PYPYPGFDRDLRWR TGVNIVPIHCA 180
Chinese    AKFDPARVVLTAGATAANELLT FIIADPGDALLVPT PYPYPGFDRDLRWR TGVNIVPIHCE 180
Apricot    AKFDPDRIVLTAGATAANELLT FIIADPGDALLVPT PYPYPGFDRDLRWR TGVNIVPIHCD 174
Lupinus    VKFDPDRIVLTAGATAANELLT FILANPGDALLVPT PYPYPGFDRDLRWR TGVNIVPIHCD 173
Western    AKFDPDRVLTAGATAANELLT FILANPGDALLVPT PYPYPGFDRDLRWR TGAKIVPIHCD 173
Castor     AKFDSDRIVLTAGATAANELLT FILADPGDALLVPT PYPYPGFDRDLRWR TGKIVPIHCD 173
Grape      AKFDPDRVLTAGATAANELLT FILADPGDALLVPT PYPYPGFDRDLRWR TSVKIVPIHCD 174
Garden     AKFNDRVLTAGATAANELLT FILADPGDALLVPT PYPYPGFDRDLRWR TGVQIVPIHCE 173
Muskmelon  AKFDPNRVLTAGATAANELLT FILANPGDALLVPT PYPYPGFDRDLRWR TGKIVPIHCD 175

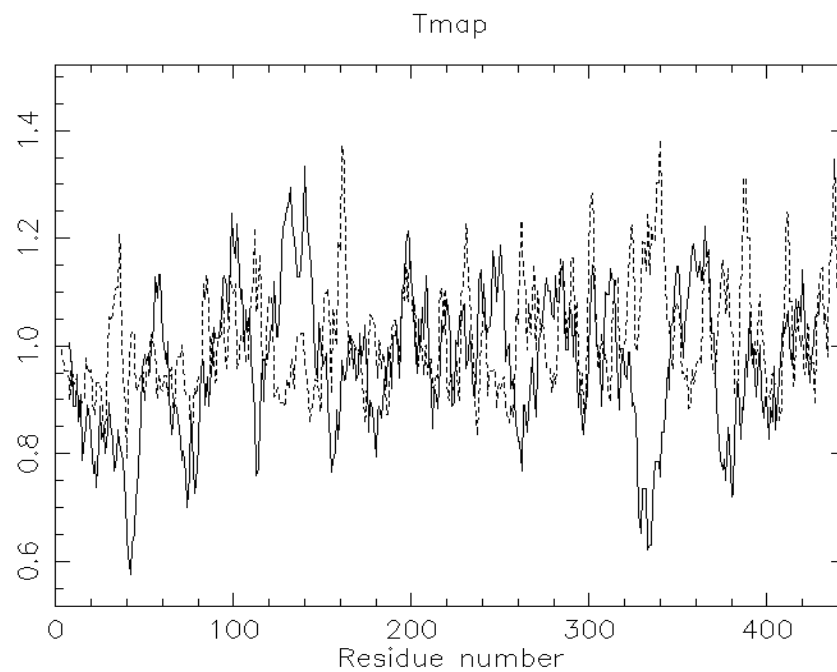
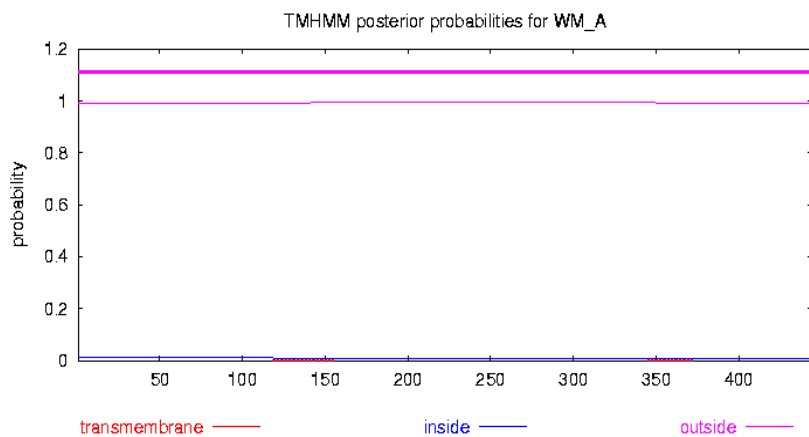
```





## TMHMM和Tmap预测蛋白质跨膜螺旋：无跨膜结构

```
# WM_A Length: 444  
# WM_A Number of predicted TMHs: 0  
# WM_A Exp number of AAs in TMHs: 0.1686  
# WM_A Exp number, first 60 AAs: 0  
# WM_A Total prob of N-in: 0.01047  
WM_A TMHMM2.0 outside 1 444
```

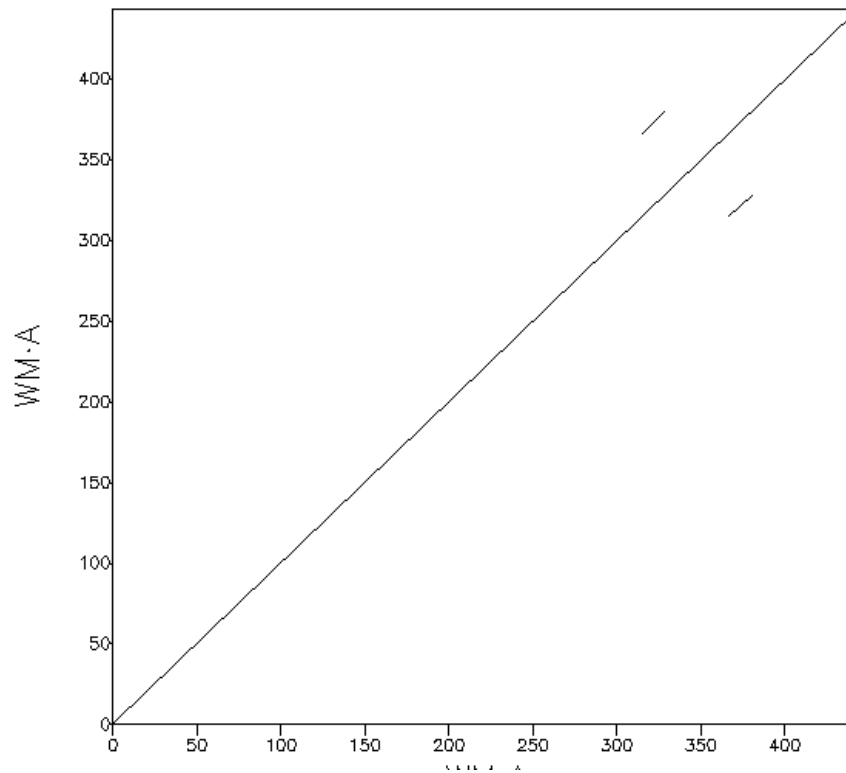






## dotmatcher分析，西瓜WA\_A蛋白有很短的一段重复序列

Dotmatcher: fasta::493411:WM-A vs fasta::493411:WM-A  
(window size = 10, threshold = 23.00 23/12/11)





## 调节雌花发育pathway

SAM(S-腺苷甲硫氨酸)



**ACS: 1-氨基环丙烷-1-羧酸合成酶**  
依赖**PLP**,  
被**AVG** (氨基乙氧基乙烯甘氨酸) 抑制

ACC(1-氨基环丙烷-1-羧酸)



**ACC氧化酶**

乙烯



促进雌花分化



注：花发育是以两性花为基础向雄花和雌花分化

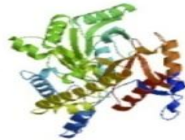
# AA, aa基因蛋白三维结构预测



Workunit: P000009 Title: WM-A

444

## Model Summary:



### Model information:

Modelled residue range: 28 to 434  
 Based on template: 1b8gA (2.37 Å)  
 Sequence Identity [%]: 60  
 Evalue: 4.01919e-149

**Quaternary structure information:**  
 Template (1b8g): DIMER  
 Model built: SINGLE CHAIN

### Ligand information:

Ligands in the template: PLP: 1.  
 Ligands in the model: none.

### Quality information:

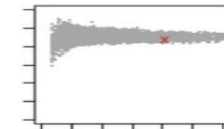
QMEAN Z-Score: -1.236

## Global Model Quality Estimation:

### QMEAN4 global scores:

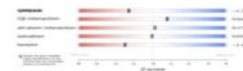
QMEANscore4: Estimated absolute model quality:

0.691



Z-Score: -1.236

### Score components:

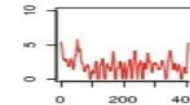


### Local scores:

Coloring by residue error:



Residue error plot:



### QMEAN4 global scores:

The QMEAN4 score is a composite score consisting of a linear combination of 4 statistical potential terms (estimated model reliability between 0-1). The pseudo-energies of the contributing terms are given below together with their Z-scores with respect to scores obtained for high-resolution experimental structures of similar size solved by X-ray crystallography:

Scoring function term	Raw score	Z-score
C_beta interaction energy	-220.13	0.84
All-atom pairwise energy	-12345.98	0.17
Solvation energy	-39.38	0.01
Torsion angle energy	-78.77	-1.46
QMEAN4 score	0.691	-1.24

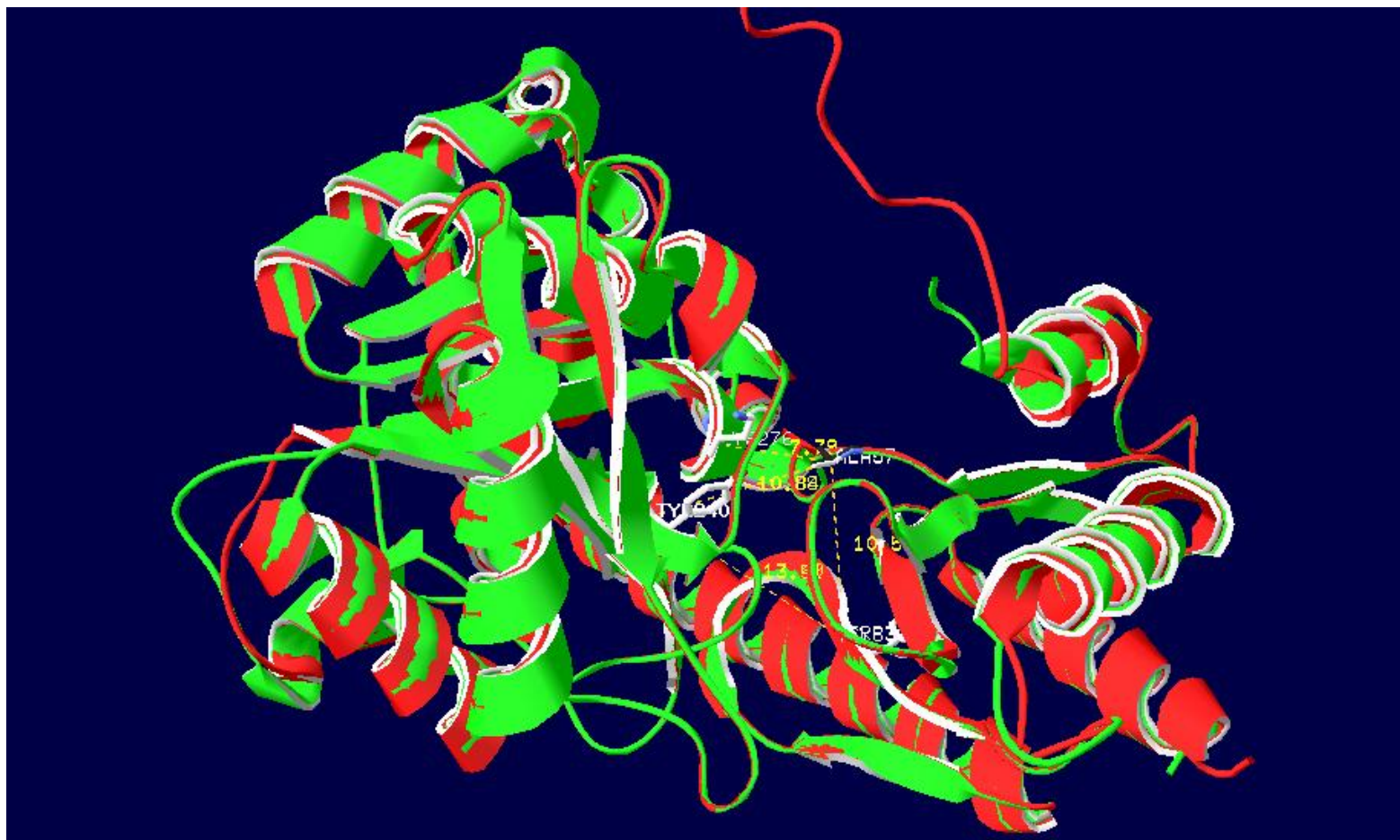
If you publish results from QMEAN, please cite the following paper:  
 Benkert P, Biasini M, Schwede T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics*, 27(3):343-50.

## Local Model Quality Estimation:





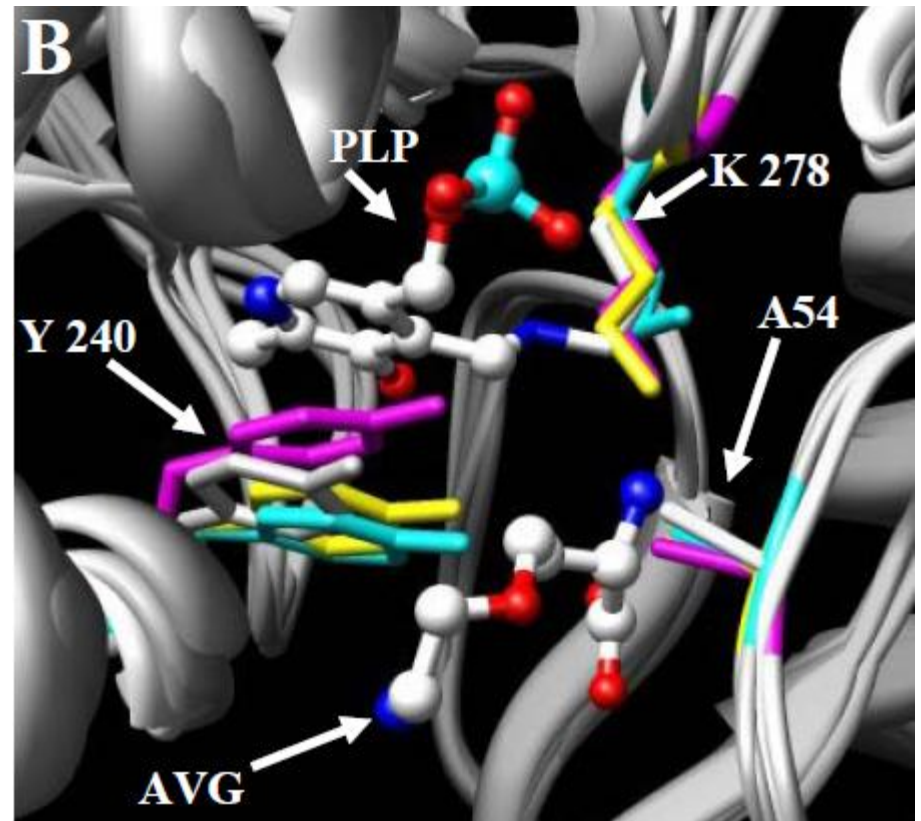
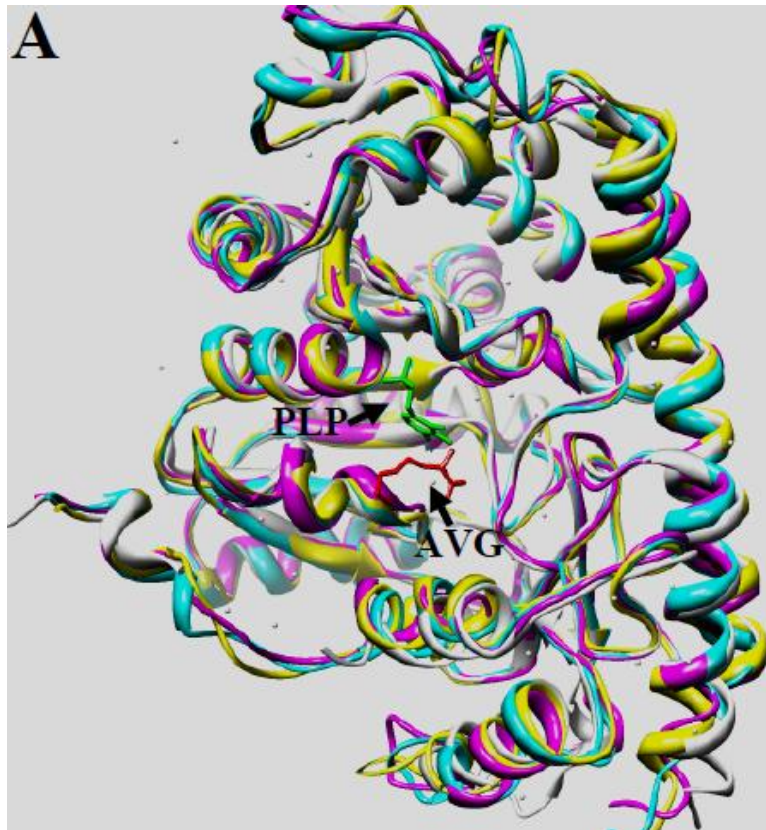
# AA, aa基因蛋白三维结构







# 甜瓜 *aa* 基因蛋白活性位点

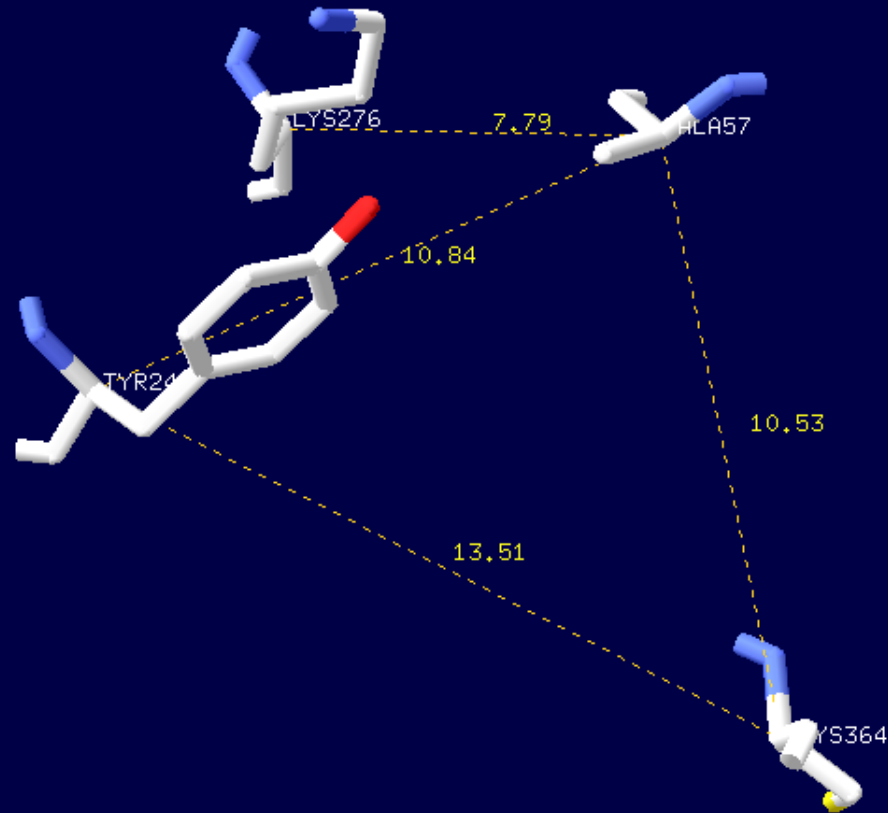


Adnane Boualem et al 2008 Science.



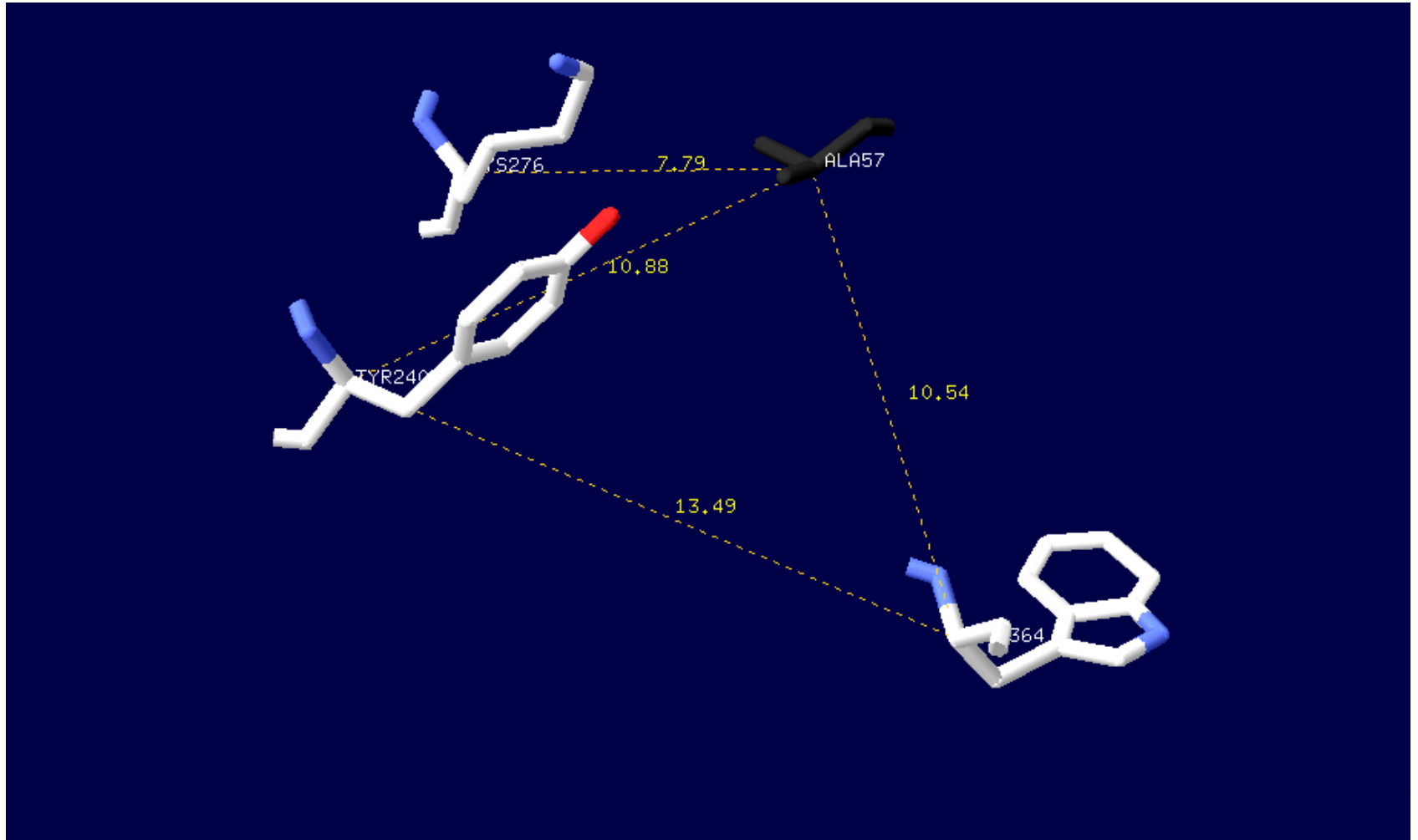


# 推测西瓜aa基因蛋白活性位点





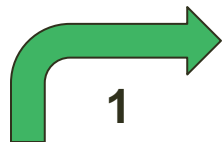
# aa活性位点氨基酸（364: C半胱氨酸-W色氨酸）



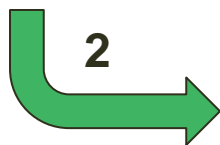


# 作用机制推测

活性位点发生  
半胱氨酸-->色氨酸突变



半胱氨酸可能是二硫键形成原因，突变后该处空间构象发生改变，导致**PLP**无法正常进入和识别。空间结构改变是无法预测出来的。



色氨酸含有苯环，较大的侧链可能  
**A**，阻止空间结构较大的激发子**PLP**进入，  
**B**，不影响空间结构小的抑制剂**AVG**进入。



**LOGO**



**Thank You !**

祝大家学习工作顺利，  
圣诞快乐！