

果胶乙酰酯酶蛋白结构分析

Structural analysis of Pectin acetylerase

3G03组：李晓飞、罗媛媛、刘亚龙、田小苗

汇报人：罗媛媛

2022-05-15



目录

CONTENTS

01. 研究背景
02. 已进行实验
03. 序列的选择
04. 理化性质分析
05. 蛋白间互作
06. 结构分析

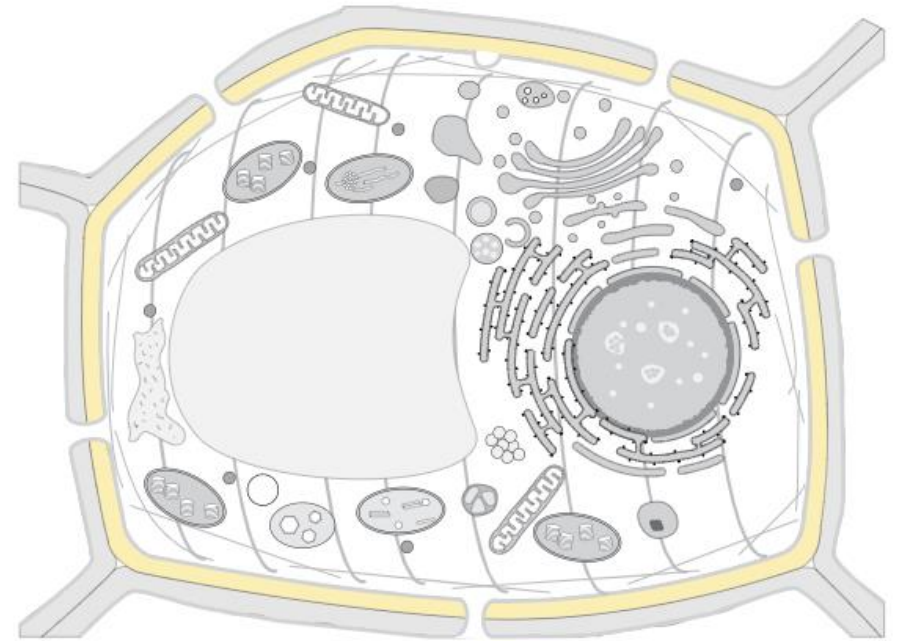


1、研究背景

- 葡萄是中国第五大经济水果，也是世界最古老的植物品种，风味独特，果肉厚且汁多味美，果实可食用率达到90%以上，有较高的经济价值。无核葡萄因其口感爽脆，更易食用，逐渐成为市场热门，而培育优良的无核葡萄品种的科研工作也正在积极开展。
- 在葡萄资源课题组的前期工作中通过蛋白质组对有核和无核葡萄果实的比较分析，鉴定到一个与种子发育相关的基因 *DUF642*，进一步研究鉴定到一个与 *DUF642* 蛋白存在相互作用的蛋白果胶乙酰酯酶 PAE (Pectin acetyl esterase)。
- 已有研究表明 PAE 能够修饰植物细胞壁结构，影响植物细胞壁稳定性、或造成部分细胞死亡、参与种子发育等。

PAE蛋白介绍

- 果胶乙酰酯酶 (Pectin acetylerase, PAE) 是果胶酯酶的一种。果胶酯酶包含果胶甲酯酶和果胶乙酰酯酶等, 果胶甲酯酶 (PME) 作用于多聚半乳糖醛酸的半乳糖醛酸残基的C-6羧基基团, 去掉甲酯, 催化果胶酯酸转化为果胶酸。果胶乙酰酯酶作用于半乳糖醛酸的C-2或C-3羧基基团, 使之乙酰化。脱乙酰作用下, 影响果胶溶解度, 使之更溶于水而降低果胶主链的疏水性。目前对乙酰化作用的研究较少, 但已证实乙酰化作用影响果胶萃取后凝胶的形成。
- 植物细胞壁通常由多糖(纤维素、半纤维素和果胶)和结构蛋白组成, 其中果胶是主要组分。果胶乙酰化发生在果胶从高尔基体向细胞壁分泌的过程中, 然后乙酰化的果胶结合进细胞壁。果胶乙酰化改变了果胶组分的理化性质, 影响了细胞的粘连进而影响了细胞壁的结构。



2.已进行的实验

- 从无核白葡萄中克隆VvPAE基因，通过农杆菌介导的遗传转化获得转VvPAE基因番茄 (*solanum* Micro-Tom)。
- 提取基因组DNA,设计PAE基因特异性引物，通过PCR鉴定基因组中是否存在VvPAE基因DNA插入。
- 提取RNA，反转录成cDNA，通过qRT-PCR分析其转录表达水平。
- 植物生长室中培养番茄植株，观察其生长发育变化，记录单果重量、种子数目及种子败育等性状。
- 通过对转基因番茄和野生型番茄基因和表型的一系列对比显示，PAE基因高峰度表达的转基因番茄在种子数量上明显减少，由此推测葡萄VvPAE基因可能影响种子发育，进而参与葡萄无核发育过程。



◆ 课题相关信息：中文名：葡萄 拉丁名：*Vitis vinifera L.*
 中文名：拟南芥 拉丁名：*Arabidopsis thaliana (Mouse-ear cress)*

登录号	登录名	蛋白名	基因名	物种
Q66GM8	PAE10_ARATH	Pectin acetylerase 10	PAE10 At5g26670, F21E10.11	Arabidopsis thaliana (Mouse-ear cress)
Q9SR23	PAE4_ARATH	Pectin acetylerase 4	PAE4 At3g09405, F3L24.30	Arabidopsis thaliana (Mouse-ear cress)
Q84JS1	PAE6_ARATH	Pectin acetylerase 6	PAE6 At3g62060, T17J13.20	Arabidopsis thaliana (Mouse-ear cress)
Q6DBP4	PAE8_ARATH	Pectin acetylerase 8	PAE8 At4g19420	Arabidopsis thaliana (Mouse-ear cress)
F4I839	PAE2_ARATH	Pectin acetylerase 2	PAE2 At1g57590, T8L23.6	Arabidopsis thaliana (Mouse-ear cress)
Q9FH82	PAE11_ARATH	Pectin acetylerase 11	PAE11 At5g45280, K9E15.6	Arabidopsis thaliana (Mouse-ear cress)
B9DFR3	PAE9_ARATH	Pectin acetylerase 9	PAE9 At5g23870, MRO11.9	Arabidopsis thaliana (Mouse-ear cress)
Q9SFF6	PAE12_ARATH	Pectin acetylerase 12	PAE12 At3g05910, F10A16.21, F2O10.13	Arabidopsis thaliana (Mouse-ear cress)
F4I107	PAE1_ARATH	Pectin acetylerase 1	PAE1 At1g09550, F14J9.21	Arabidopsis thaliana (Mouse-ear cress)
Q80731	PAE3_ARATH	Pectin acetylerase 3	PAE3 At2g46930	Arabidopsis thaliana (Mouse-ear cress)
Q9SR22	PAE5_ARATH	Pectin acetylerase 5	PAE5 At3g09410, F3L24.31	Arabidopsis thaliana (Mouse-ear cress)
Q940J8	PAE7_ARATH	Pectin acetylerase 7	PAE7 At4g19410, T5K18.190	Arabidopsis thaliana (Mouse-ear cress)
A0A438HSE5	A0A438HSE5_VITVI	Pectin acetylerase	PAE5_1 CK203_033640	Vitis vinifera (Grape)

3、序列的选择

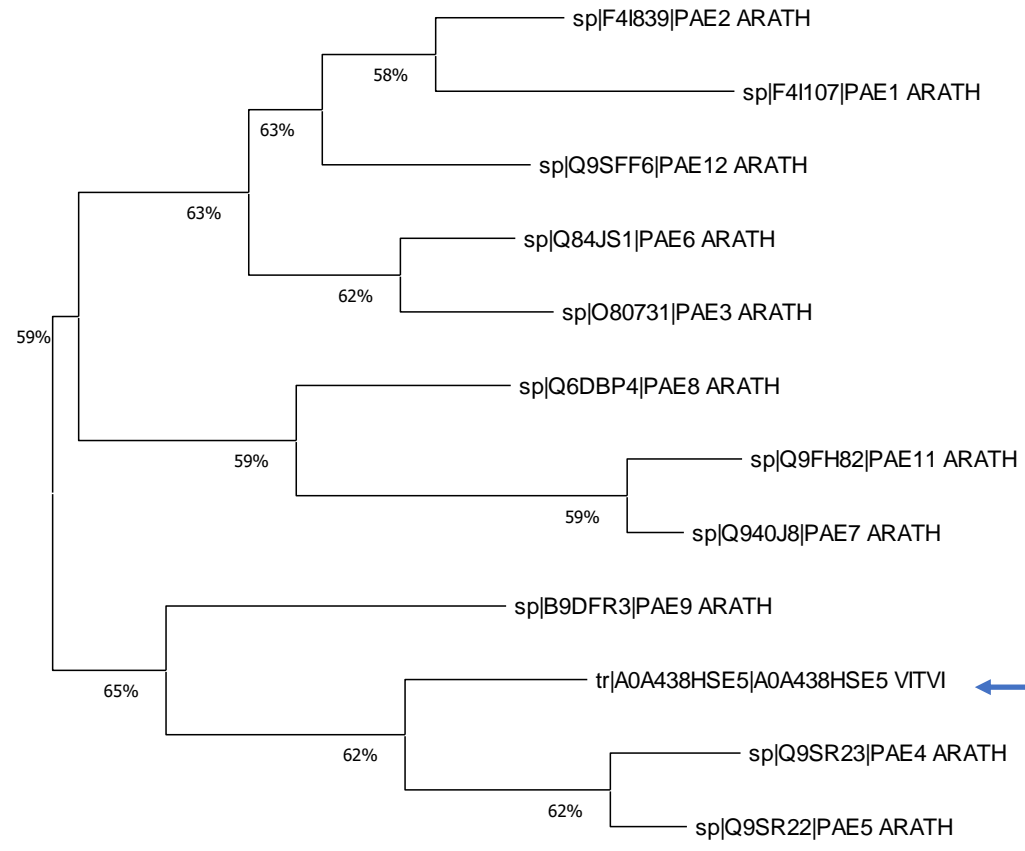
分析过程中存在的问题：
在UniProt关于葡萄PAE蛋白
只有一个未经人工审核的蛋
白序列，准确性得不到保证

解 决 问 题 思 路

1、构建mega系统发生树，将葡萄和拟南芥的PAE蛋白序列进行建树，找到结构最相似的。

2、通过NCBI-Blast查询相似度最高的序列。

3、通过Swiss-model找到与A0A438HSE5_VITVI最相近的结构。



← 葡萄PAE

} 拟南芥PAE4、5

0.20

在进行葡萄PAB蛋白NCBI-Blast时发现
A0A438HSE5_VITVI在263位氨基酸出现：X-未知
氨基酸，无法使用

blast.ncbi.nlm.nih.gov 显示

Your organism was not found in the taxonomic lookup. Please,
select a valid organism from the suggestions

确定

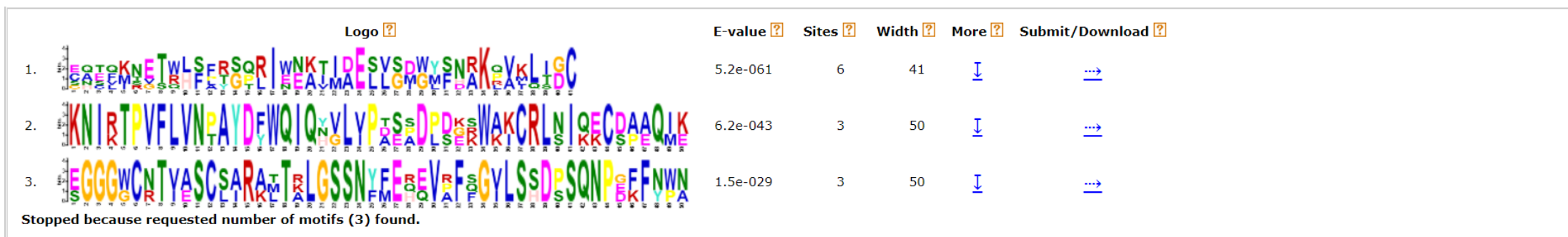


10	20	30	40	50
MALFLSLCLV	FFSIVCSSAS	ATLPPYLDGL	LPNGDFEKG	KASNLKKTVI
60	70	80	90	100
KGKHSPLDWE	IKGLVEYVSG	GPQPGGFFLA	VPRGWHAVKL	GNEASISQNV
110	120	130	140	150
RVRPGFIYSL	TFGATRTCAQ	DEVLRVSIPG	QSADLPIQTL	YSTDGGDTYA
160	170	180	190	200
WAFNATTEVW	NVTFHNPQIQ	EDRTCGPLLD	AIAIKQMPPL	KLTRGSLAKN
210	220	230	240	250
GGYETGPHVF	KMFSTGVLLL	PKQQDLISPL	PGWIIESLKP	VKYIDSKHFS
260	270	280	290	300
VPRGLAAIEL	IGGRESAIAQ	IIRTVPNKSY	NLTFTIGDAK	NGCHGSMTVE
310	320	330	340	350
AFAAKETVKV	PHASQKGGEF	KAASLKFKAL	SARTRITFYS	AYYHTKLHDF
360	370	380	390	400
GHMCGPVLLD	VRMSRSGALD	LASGVGGKIE	KNEVLSAVQK	YEKYHVCHGG
410	420	430	440	450
DEEERRANYT	DMVNKYYDLA	TSFYEYGWGE	SFHFASRWKG	ESLRESIKRH
460	470	480	490	500
EHFLAVQLGL	KCGQKVLVDG	CGIGGPLREI	ARFSLGALVA	LQELNRLAGV
510	520	530	540	550
DKTCNFELFL	LMDNQADFMK	MPFSNDTFDA	VFAIEATCHA	PDVLDYKEI
560	570	580	590	600
YRVLPKPGQCF	AAAYEWCITDC	FDPMNREHQR	IKGEVELGNG	LPDIRSVGQC
610	620	630	640	650
LEALKLAGFE	VLWEKDVAAD	SPLPWYLPD	TTQFSLSNFR	TTSFGRFITR
660	670	680	690	700
NMVKALEFVR	LAPEGSQRVQ	AFLEKAAEAL	VEGGREGIFT	PMYFFVARKP

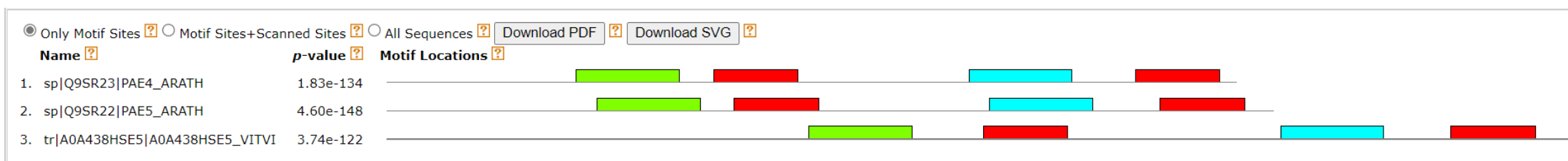
同样在进行Swiss-model时也出现同样问题
通过参考系统发生树演化关系，后续关于
PAE的分析采用拟南芥PAE4、5蛋白

保守性分析

DISCOVERED MOTIFS



MOTIF LOCATIONS

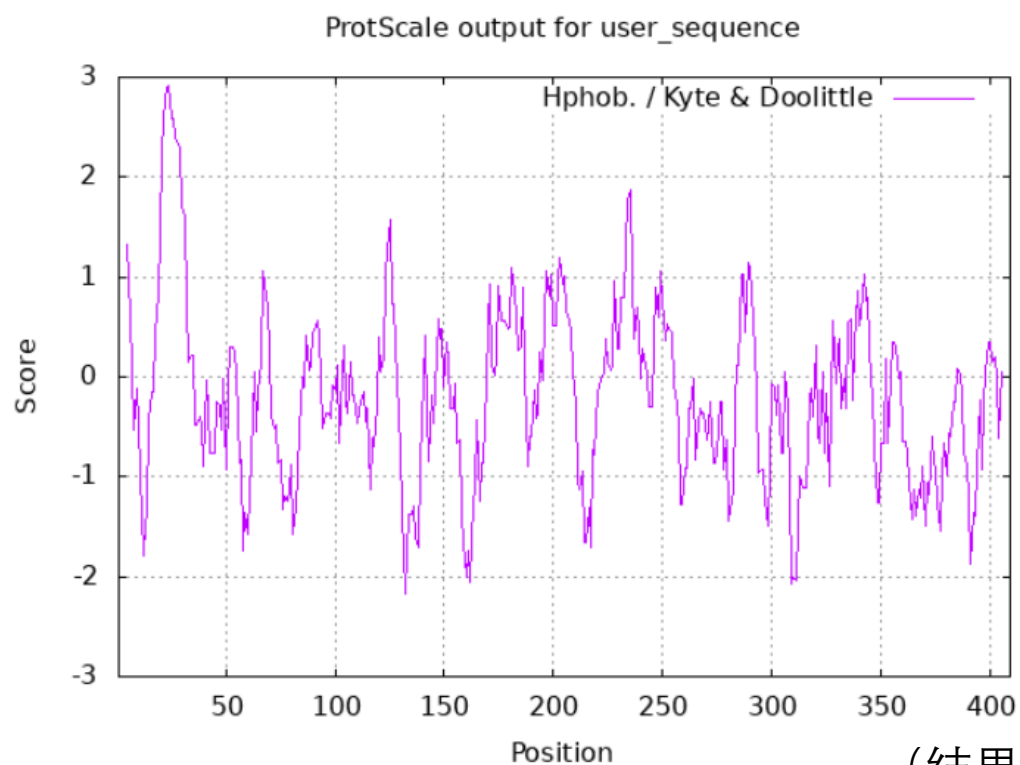


(结果来自MEME)

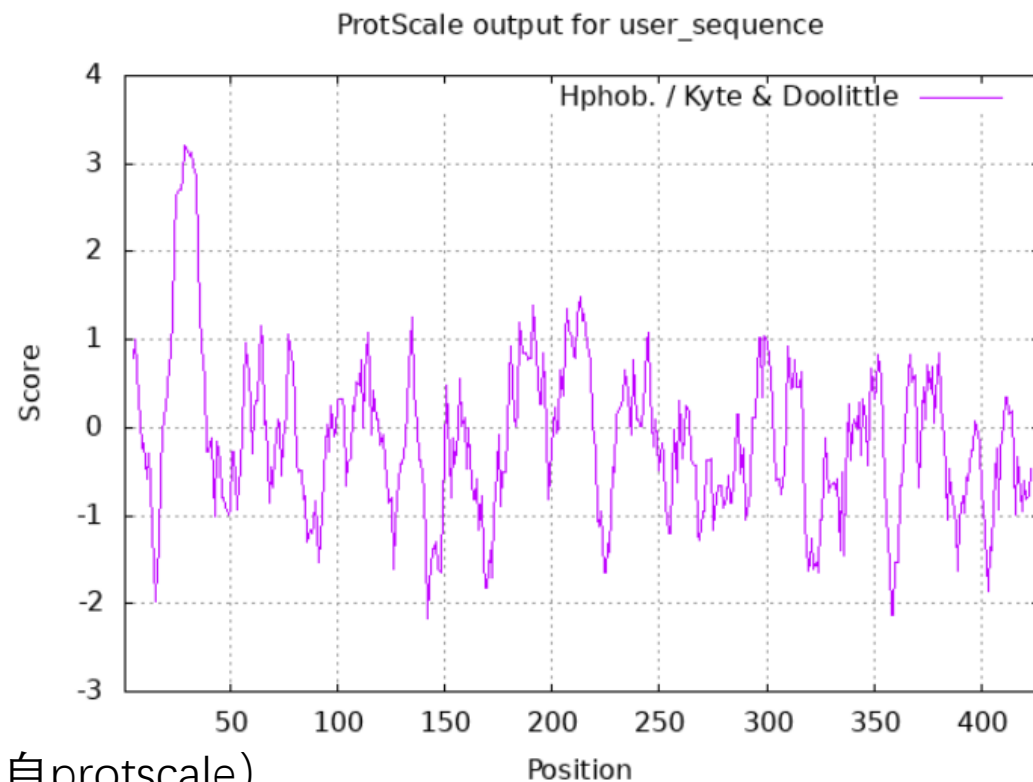
4、理化性质

(结果来自protparam)

subunit	accession	Number of amino acids	Molecular weight	Theoretical pI	Asp+Glu	Arg+Lys	instability index	GRAVY
PAE4_ARATH	Q9SR23	409	45749.12	6.31	44	40	41.26	-0.174
PAE5_ARATH	Q9SR22	427	47637.42	6.81	43	42	42.74	-0.137

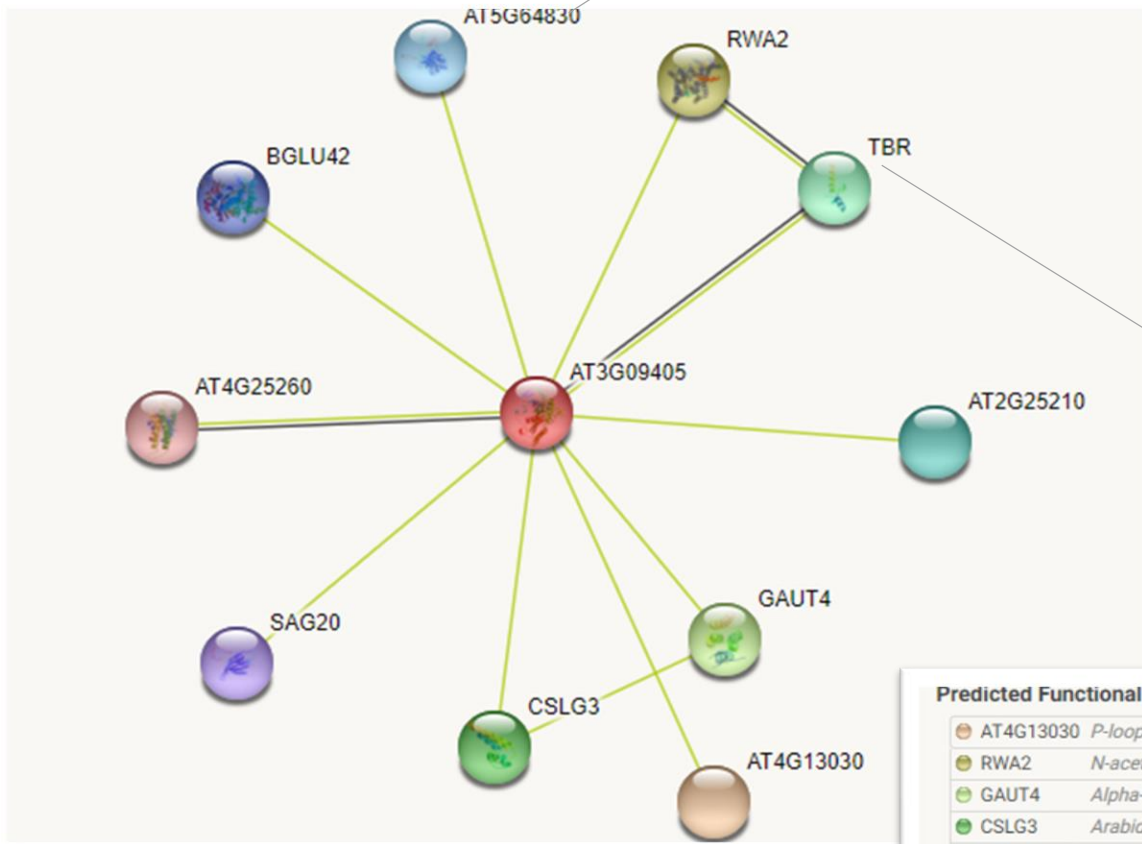


(结果来自protscale)



5、蛋白互作 PAE4_ARATH

AT5G64830：程序性细胞死亡 2 C-末端结构域蛋白；参与细胞凋亡；位于细胞质中；以 22 种植物结构表达；在 13 个生长阶段表达；



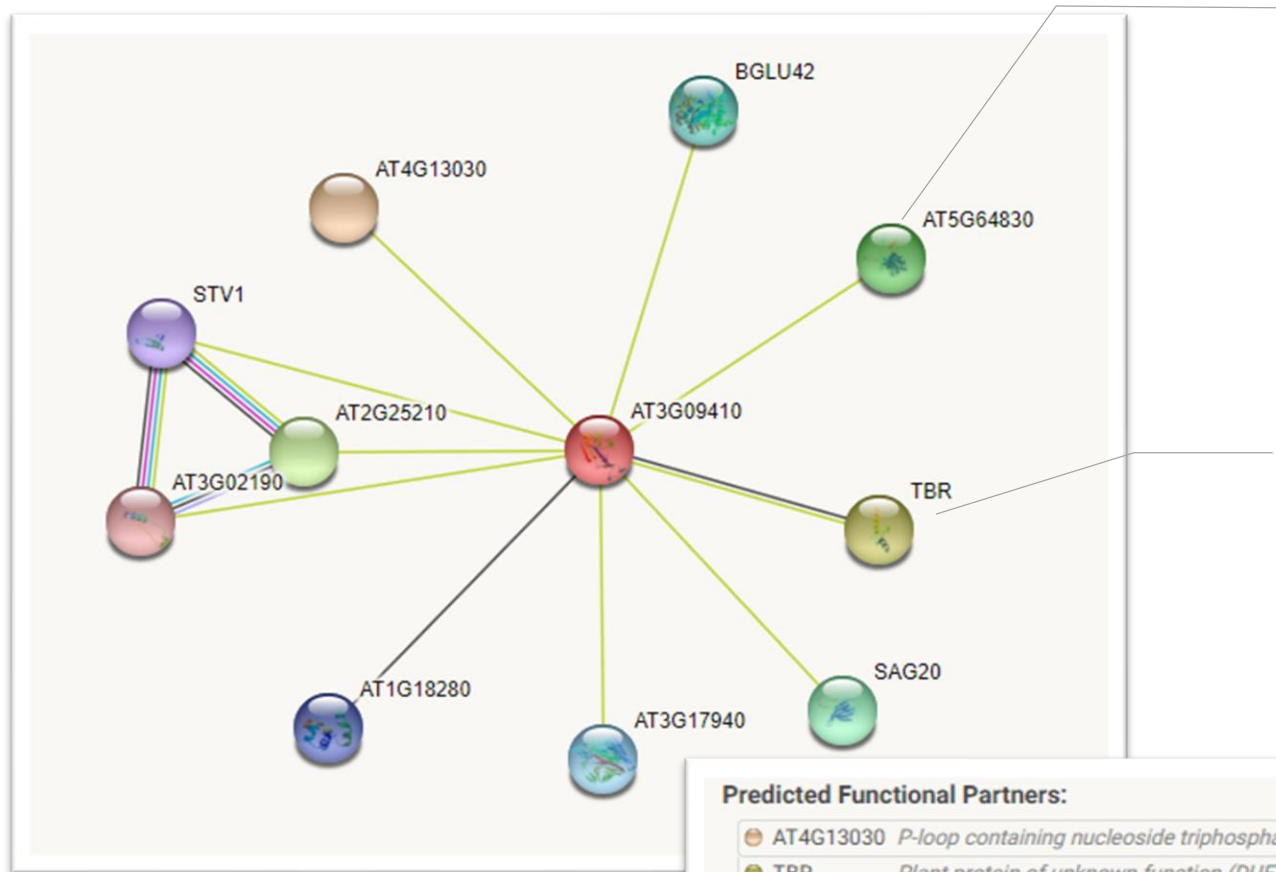
未知功能的植物蛋白 (DUF828)；编码 TBL 基因家族的成员，该家族包含植物特异性 DUF231 (未知功能域) 域。TBL 基因家族有 46 个成员，其中两个 (TBR/AT5G06700 和 TBL3/AT5G01360) 已被证明参与次生壁纤维素的合成和沉积，可能是通过影响果胶聚合物的酯化状态

Predicted Functional Partners:

		Ne	Ge	Co	Ca	Ex	Da	Te	/Ht	Sc
●	AT4G13030	P-loop containing nucleoside triphosphate hydrolases superfamily protein; Has 75 Blast hits to 75 proteins in 36 spe...							●	0.805
●	RWA2	N-acetylneuraminate 9-o-acetyltransferase; Probable O-acetyltransferase involved in the acetylation of cell wall poly...							●	0.616
●	GAUT4	Alpha-1,4-galacturonosyltransferase; Probable galacturonosyltransferase 4; Encodes a protein with putative galactur...							●	0.593
●	CSLG3	Arabidopsis thaliana cellulose synthase-like g3; Cellulose synthase-like protein G3; Thought to be a Golgi-localized be...							●	0.587
●	TBR	Plant protein of unknown function (DUF828); Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene...							●	0.582
●	AT2G25210	Ribosomal protein L39 family protein; Its function is described as structural constituent of ribosome; Involved in tran...							●	0.580
●	AT5G64830	Programmed cell death 2 C-terminal domain-containing protein; Its function is described as molecular_function unkn...							●	0.580
●	BGLU42	Beta-glucosidase 42; Involved in the secretion of root-derived phenolics upon iron ions (Fe) depletion. Promotes dise...							●	0.575
●	SAG20	Arabidopsis thaliana wound-induced protein 12; A senescence-associated gene whose expression is induced in resp...							●	0.563
●	AT4G25260	Plant invertase/pectin methylesterase inhibitor superfamily protein; Pectin methylesterase (PME) inhibitor that can t...							●	0.562

(结果来自STRING)

5、蛋白互作 PAE5_ARATH



AT5G64830: 程序性细胞死亡 2 C-末端结构域蛋白; 它的功能被描述为molecular_function unknown; 参与细胞凋亡; 位于细胞质中; 以 22 种植物结构表达; 在 13 个生长阶段表达;

未知功能的植物蛋白 (DUF828); TBL基因家族有46 个成员, 其中两个 (TBR/AT5G06700和 TBL3/AT5G01360) 已被证明参与次生壁纤维素的合成和沉积, 可能是通过影响果胶聚合物的酯化状态。

(结果来自STRING)

Predicted Functional Partners:

		Nei	Ger	Coc	CoE	Exp	Dat	Tex	I/No	Scd
AT4G13030	<i>P-loop containing nucleoside triphosphate hydrolases superfamily protein; Has 75 Blast hits to 75 proteins in 36 spe...</i>								●	0.809
TBR	<i>Plant protein of unknown function (DUF828); Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene...</i>					●			●	0.587
AT2G25210	<i>Ribosomal protein L39 family protein; Its function is described as structural constituent of ribosome; Involved in tran...</i>								●	0.584
AT5G64830	<i>Programmed cell death 2 C-terminal domain-containing protein; Its function is described as molecular_function unkn...</i>								●	0.584
SAG20	<i>Arabidopsis thaliana wound-induced protein 12; A senescence-associated gene whose expression is induced in resp...</i>								●	0.581
BGLU42	<i>Beta-glucosidase 42; Involved in the secretion of root-derived phenolics upon iron ions (Fe) depletion. Promotes dise...</i>								●	0.575
AT3G17940	<i>Galactose mutarotase-like superfamily protein; Converts alpha-aldose to the beta-anomer</i>								●	0.515
AT1G18280	<i>Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; Contains the following Inter...</i>					●				0.483
STV1	<i>Ribosomal protein l24e family protein; RPL24B encodes ribosomal protein L24, homolog of cytosolic RPL24, found in...</i>								●	0.463
AT3G02190	<i>Ribosomal protein L39 family protein; Its function is described as structural constituent of ribosome; Involved in tran...</i>								●	0.461

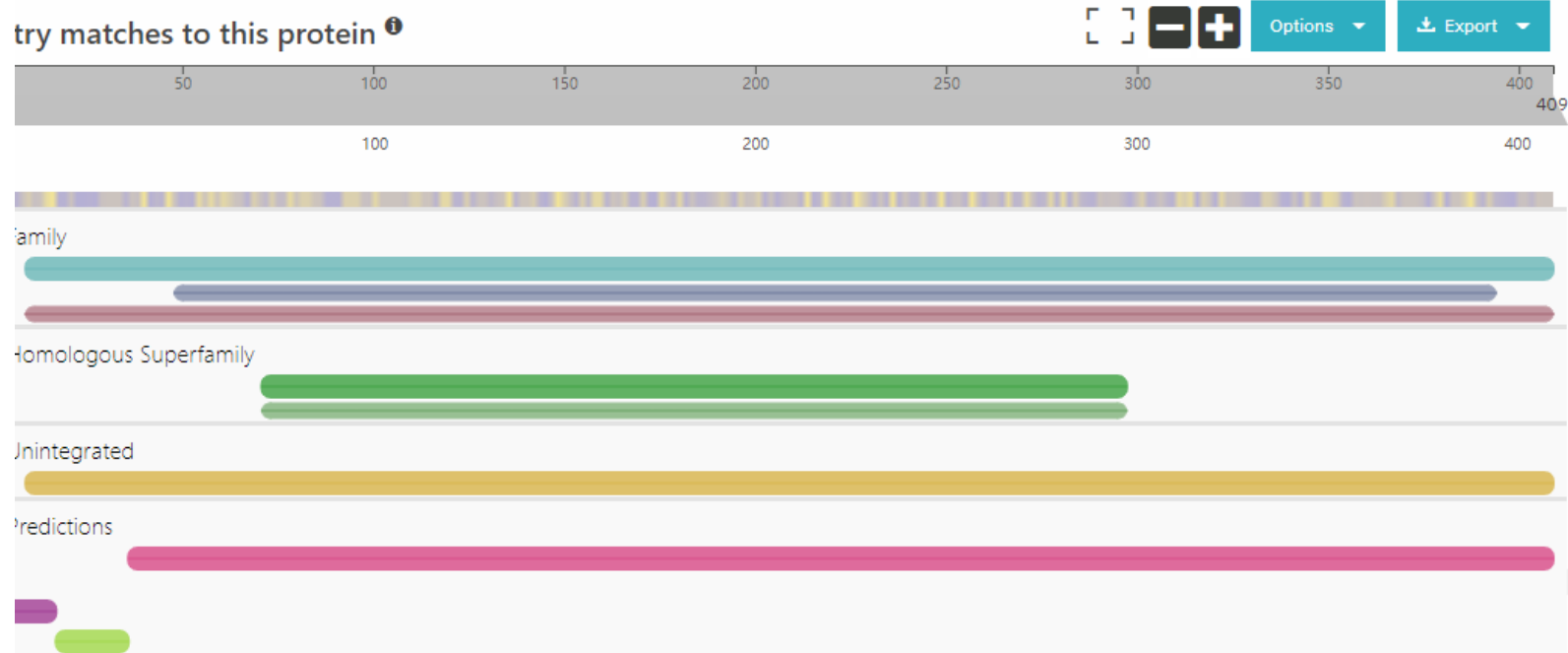
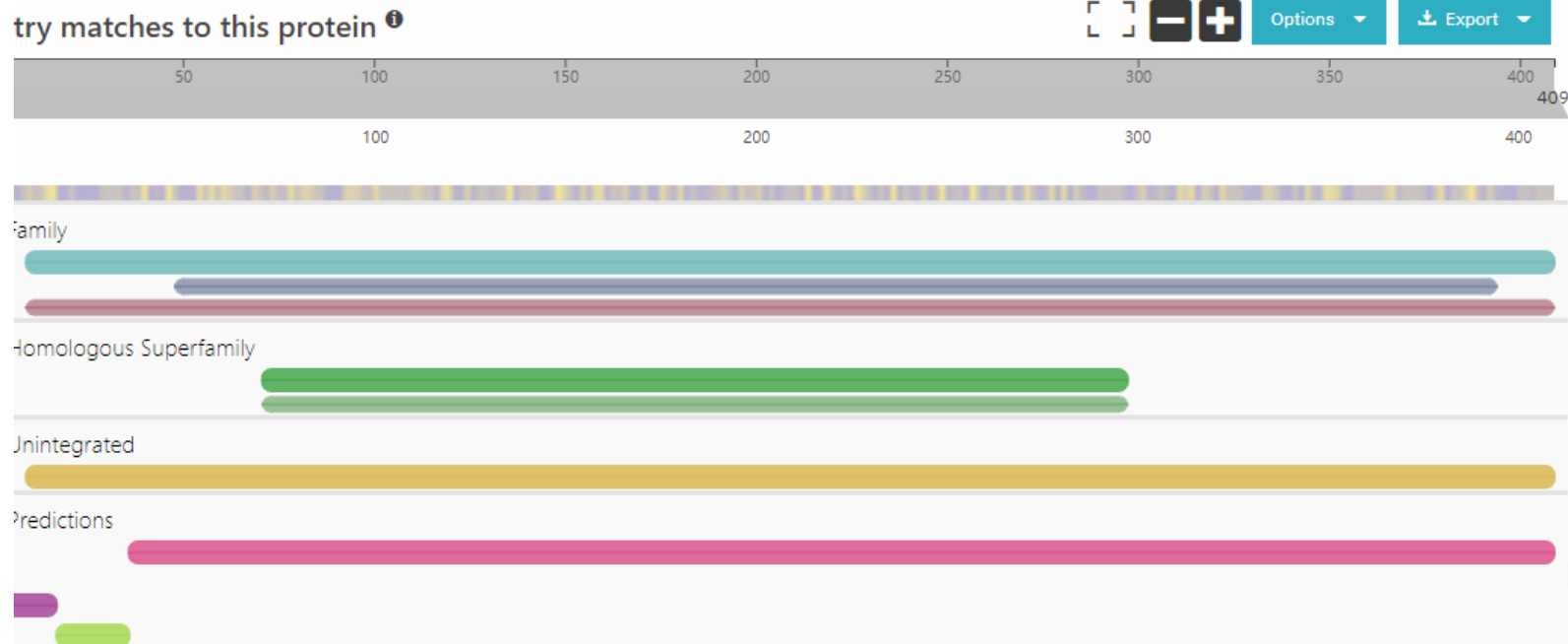
二级结构

- 上PAE_4
- 下PAE_5
- 在预测结构中发现跨膜区:

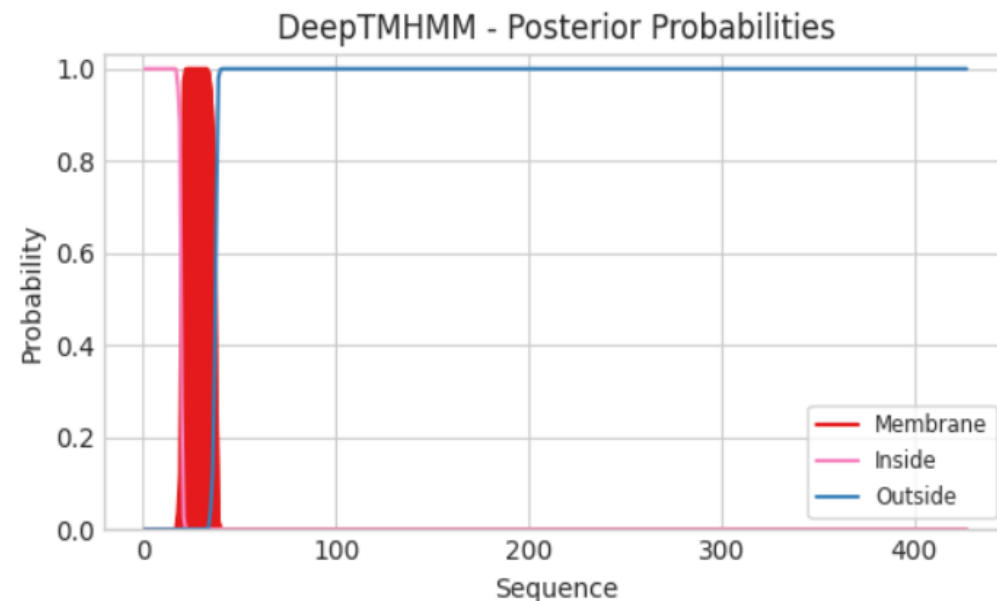
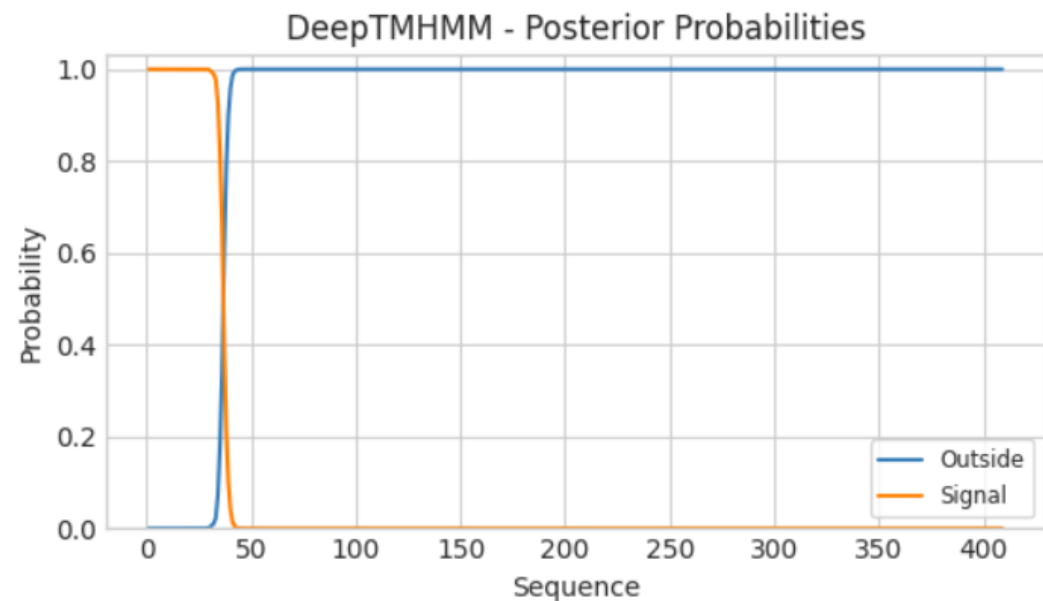
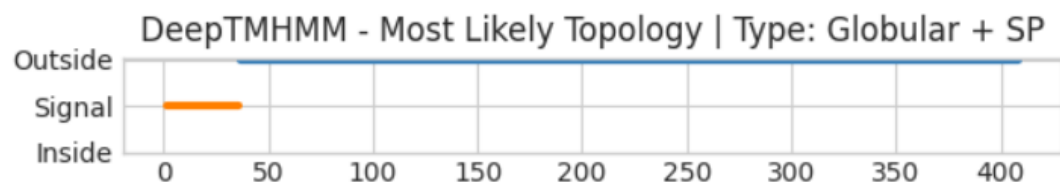
PAE_4: 17-35

PAE_5: 20-38

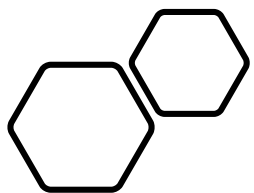
(结果来自IntroPro)



从图中看出可以signal代表信号肽, outside代表膜外, inside代表膜内, membrane为跨膜区。显然, PAE_4: 共有409个氨基酸残基, 信号肽为 1-36位氨基酸, 37-409位在膜外。PAE_5: 1-19位为膜内氨基酸, 38-427位为膜外氨基酸。存在一个跨膜区域, 位于20-37位氨基酸, 不存在信号肽。



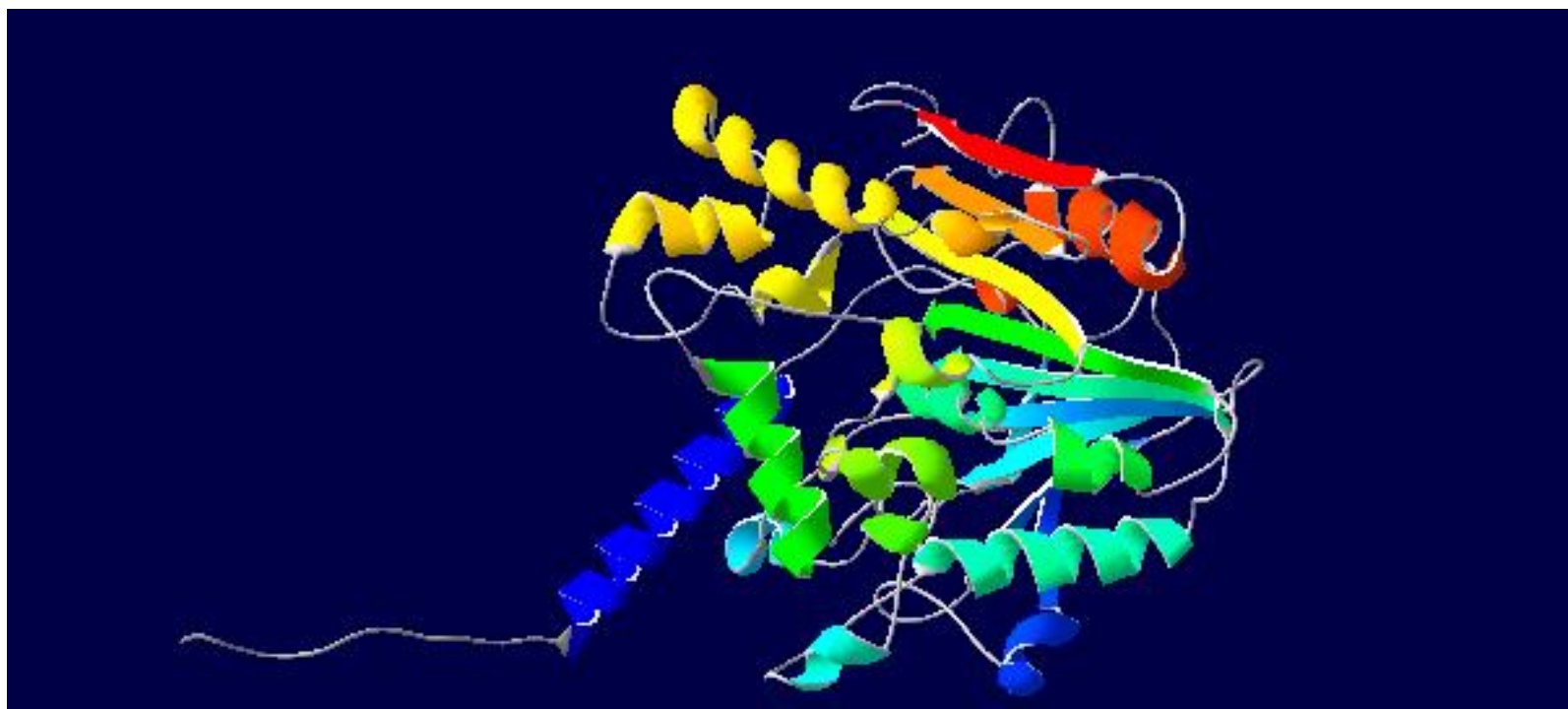
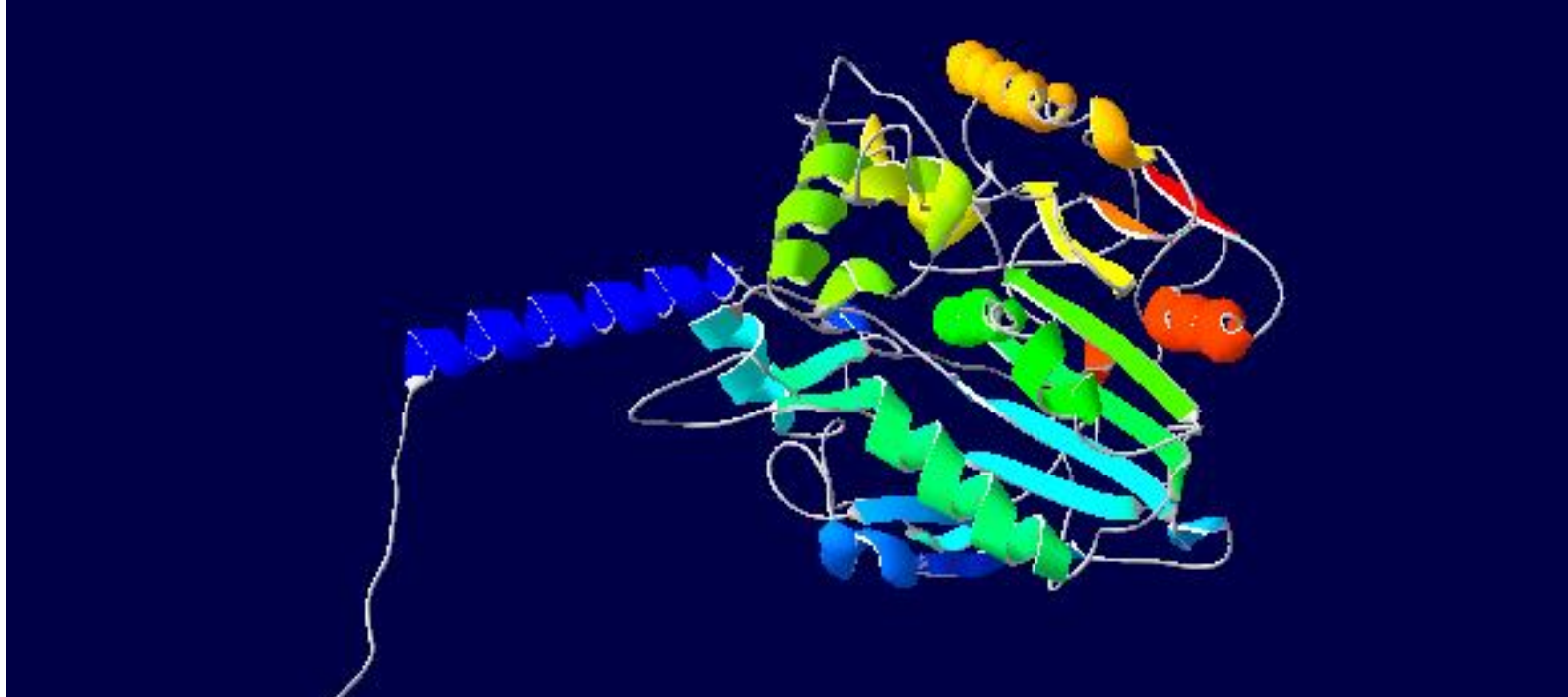
(结果来自TMHMM)



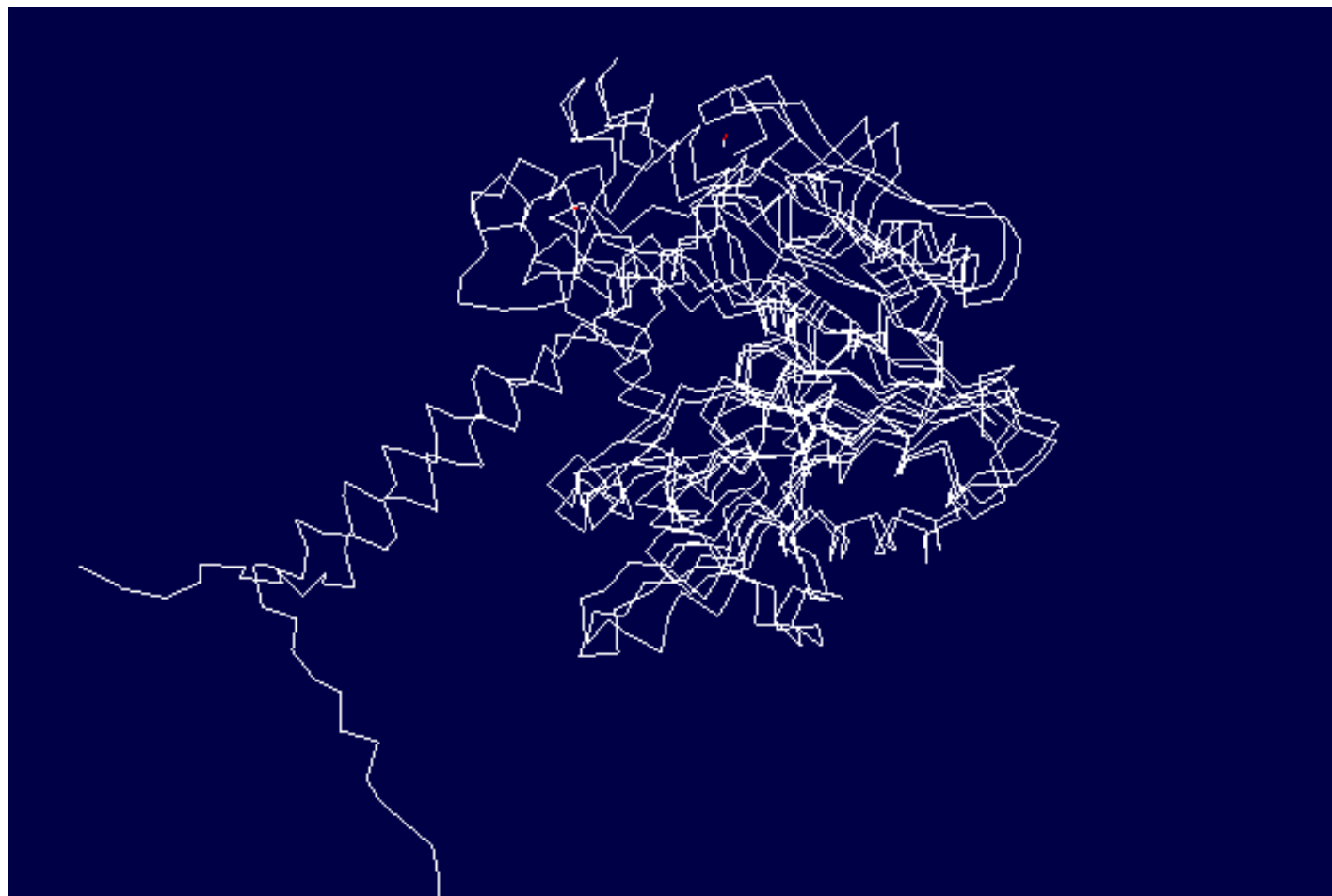
三级结构

- 上PAE_4
- 下PAE_5

(结果来自spdbv)



三维结构



(结果来自spdbv)

通过对PAE_4和PAE_5进行均方根误差分析, 得到RMSD值为6.03

总结：

- 通过对拟南芥PAE蛋白的分析，在互作蛋白中发现有和细胞程序性死亡和次生壁纤维素的合成和沉积相关蛋白，侧面证实了葡萄PAE蛋白无核葡萄之间的关系。
- 后续仍需要对无核白葡萄的PAE蛋白序列进行分析。

感谢聆听

请老师批评指正!