

Bioinformatic analysis of TSPO protein in Arabidopsis

拟南芥TSPO蛋白的生物信息学分析

Reporter: Song Zihao

Group member: Liu J, Zhang H
Zhou Q, Song Z

2014-1-25



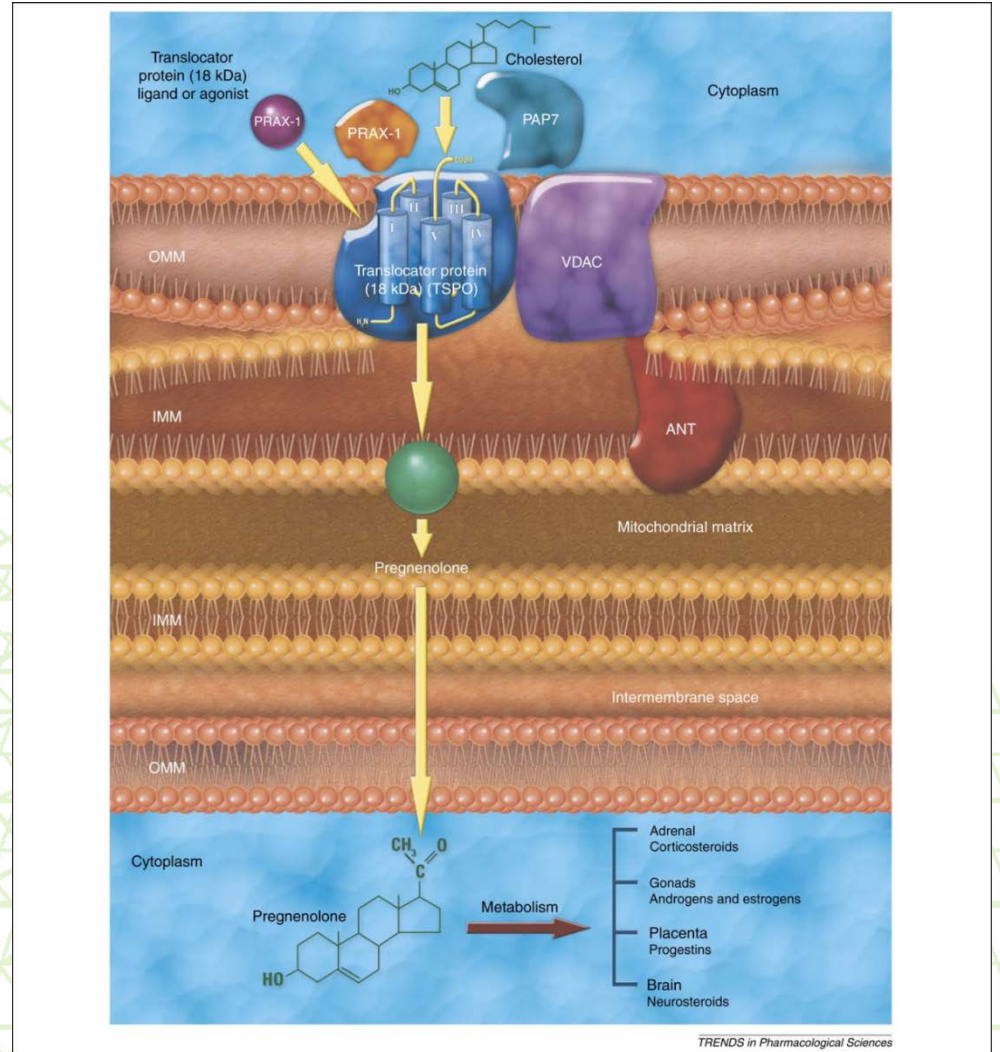
背景介绍

生物信息学分析

总结

背景介绍

- TspO/MBR蛋白家族
- 类固醇类合成
- 蛋白转运
- 卟啉衍生物的运输和亚铁血红素的合成
- 离子转运



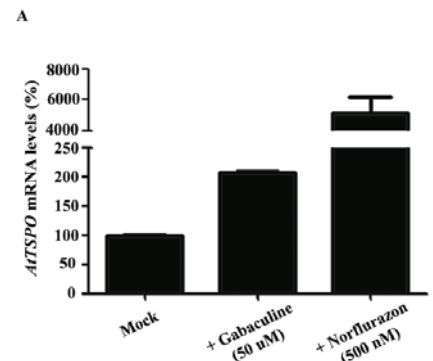
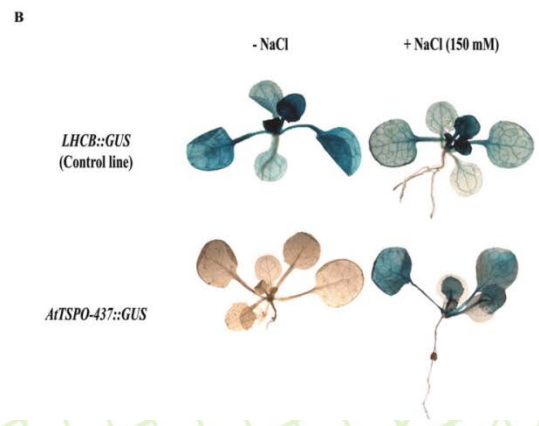
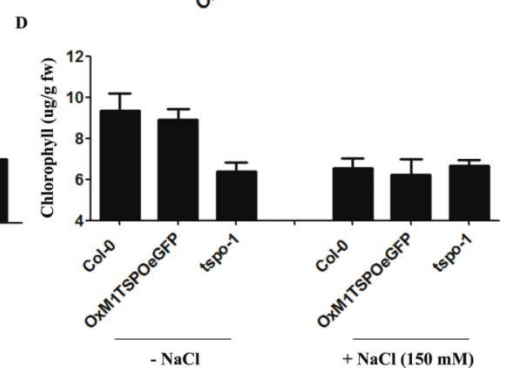
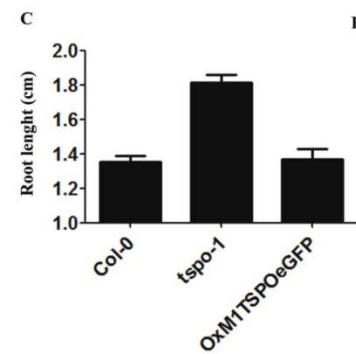
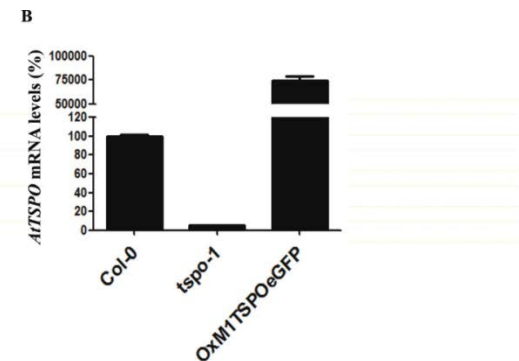
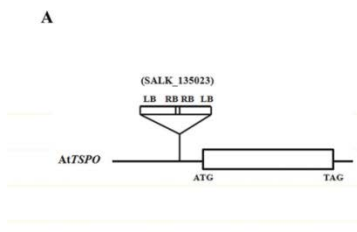
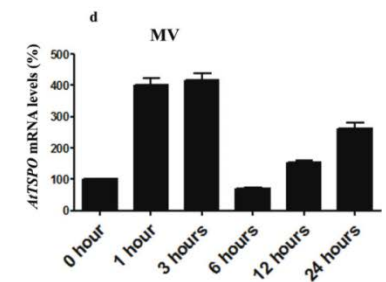
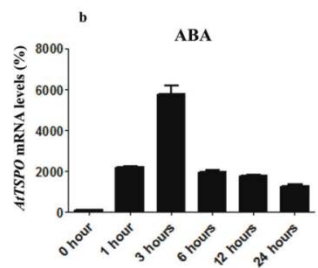
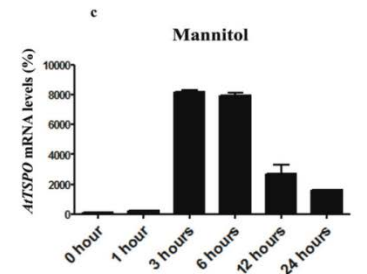
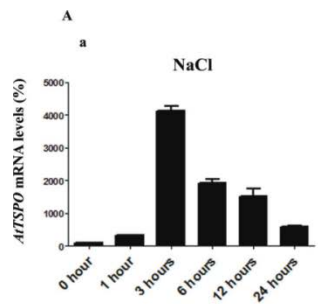
(Papadopoulos et al., Steroids, 1997)

TSPO 种间同源性分析

	Acc. No.	Length	Identity, %		
			<i>A. thaliana</i>	<i>R. capsulatus</i>	<i>B. taurus</i>
<i>Rattus norvegicus</i>	P16257	169	21.6	29.8	78.1
<i>Homo sapiens</i>	P30536	169	21.6	31.1	74.6
<i>Mus musculus</i>	Q62118	169	22.5	27.0	75.7
<i>Bos taurus</i>	P30535	169	21.8	27.1	
<i>Acidophilium rubrum</i>	BAA76530	168	17.8	30.9	31.7
<i>Heliobacillus mobilis</i>	T31451	168	18.8	25.7	25.8
<i>Rhodobacter capsulatus</i>	P17057	160	15.5		27.1
<i>Chlamydomonas reinhardtii</i>	BG858044 ^b	169	21.3	26.4	26.6
<i>Physcomitrella patens</i>	BJ159476 ^b	177	20.9	35.9	36.0
<i>Marchantia polymorpha</i>	AU081713 ^b	162	20.6	35.3	37.9
<i>Pinus taeda</i>	AI725033 ^b	168	20.1	35.2	37.8
<i>Cycas rumphii</i>	CB091062 ^b	198	28.0	38.7	33.2
<i>Solanum tuberosum</i>	BG888287 ^{a,b}	146	30.4	12.1	12.7
<i>Lycopersicon esculentum</i>	AI782627 ^{a,b}	136	32.7	23.4	24.9
<i>Medicago truncatula</i>	CA858258 ^b	200	39.8	16.6	20.7
<i>Glycine max</i>	BM891625 ^b	179	42.3	20.0	24.8
<i>Citrus sinensis</i>	CB293942 ^{a,b}	156	41.3	19.5	27.2
<i>Populus tremula</i>	S019C02 ^b	197	47.0	19.0	23.9
<i>Brassica oleracea</i>	BZ489380 ^b	195	85.4	14.5	20.7
<i>Arabidopsis thaliana</i>	O82245	196		15.5	21.8
<i>Synechocystis</i>	P74502	159	16.9	24.4	21.0
<i>Synechocystis</i>	Q55180	151	7.5	4.7	11.5

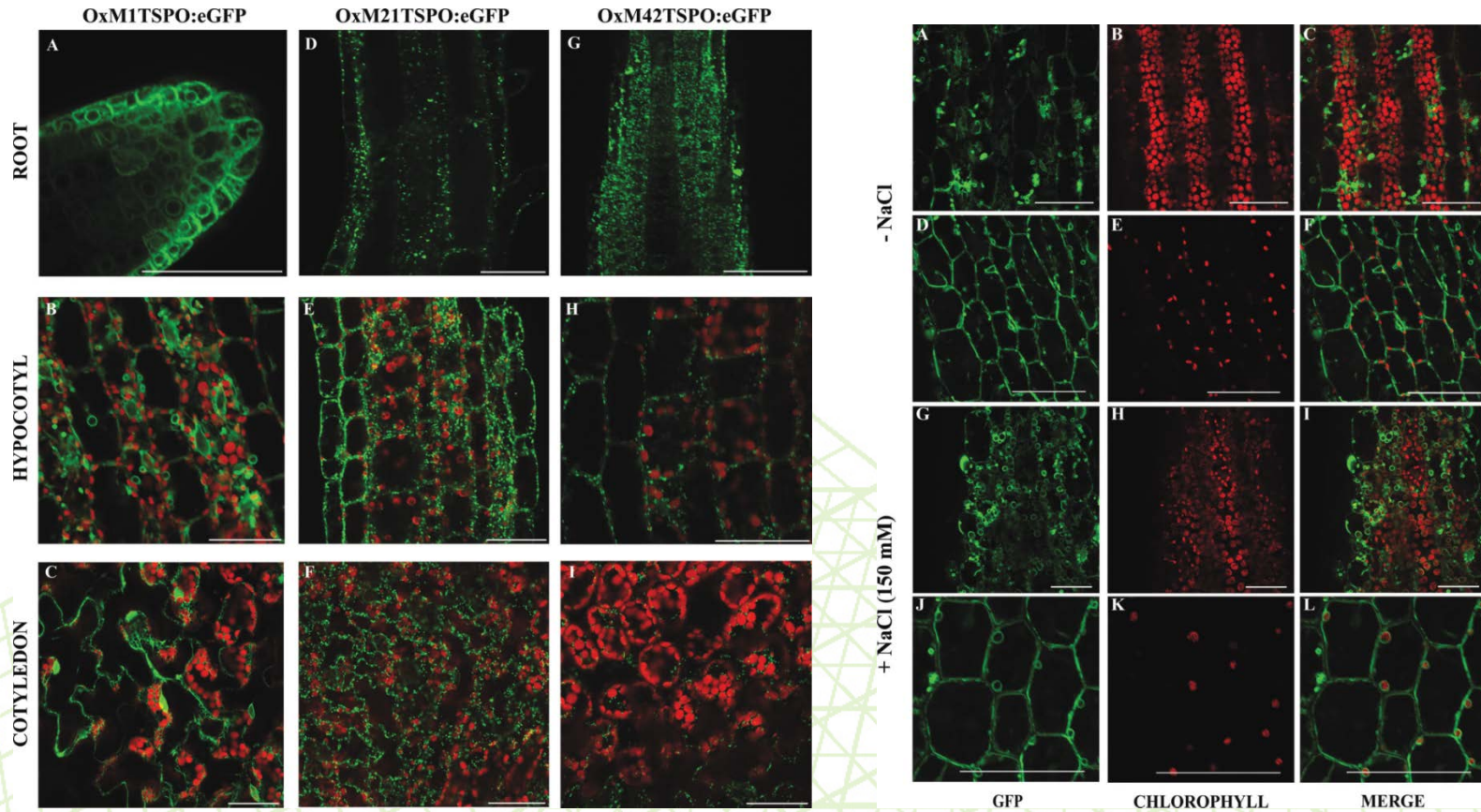
(Lindemann et al., Plant Cell Physiol, 2004)

TSPO 对逆境的响应



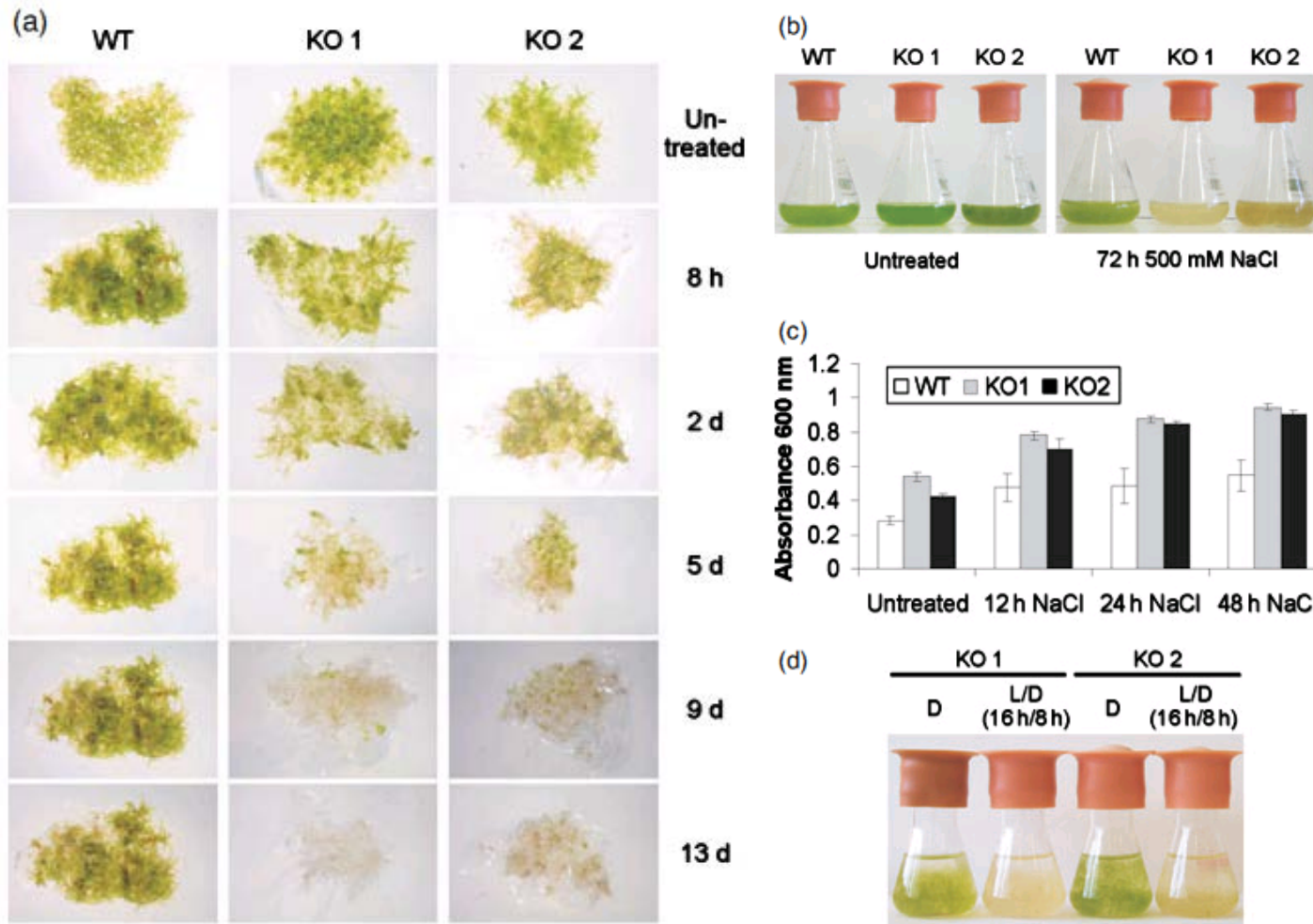
(Balsemão-Pires et al., BMC Plant Biology, 2011)

TSPO 定位

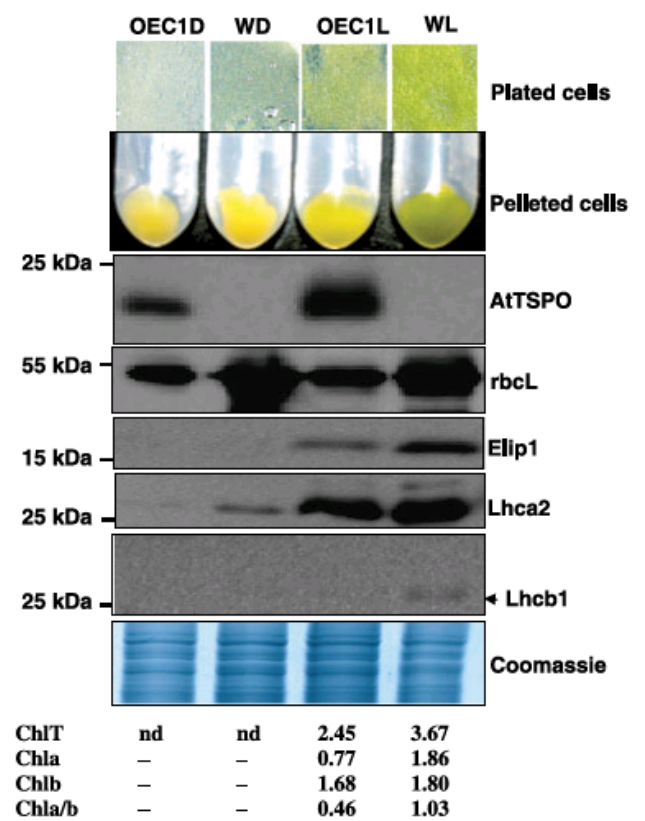
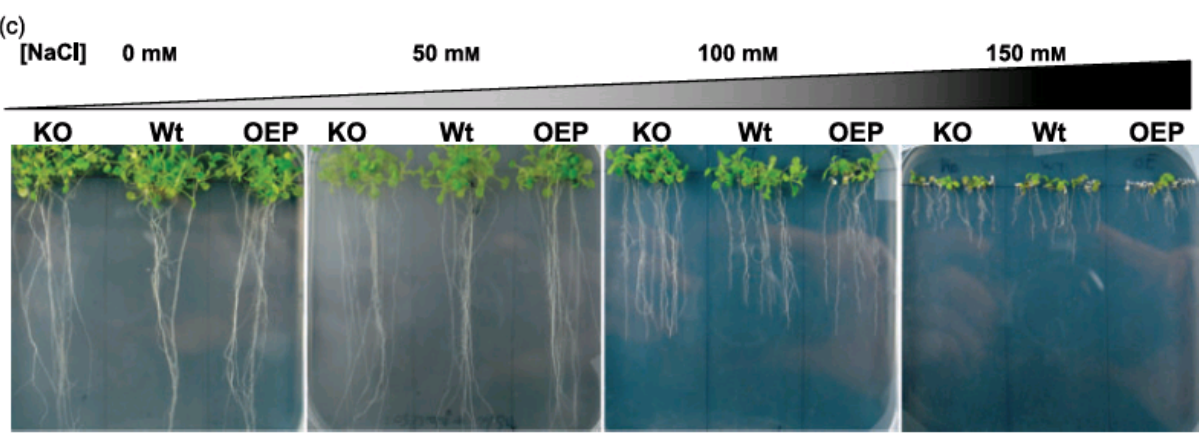
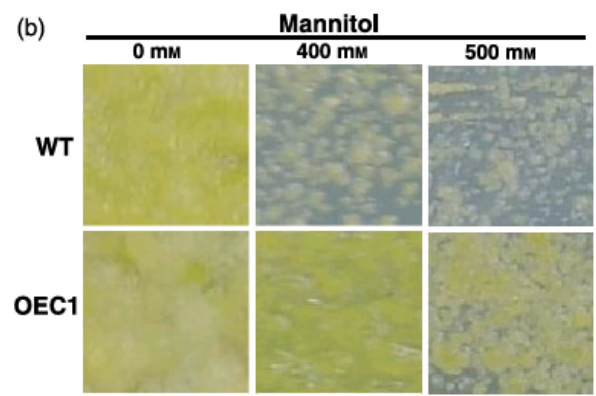
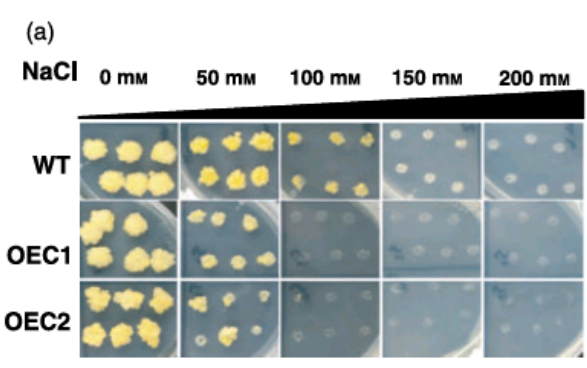


(Balsemão-Pires et al., BMC Plant Biology ,2011)

小立碗藓中TSP0研究

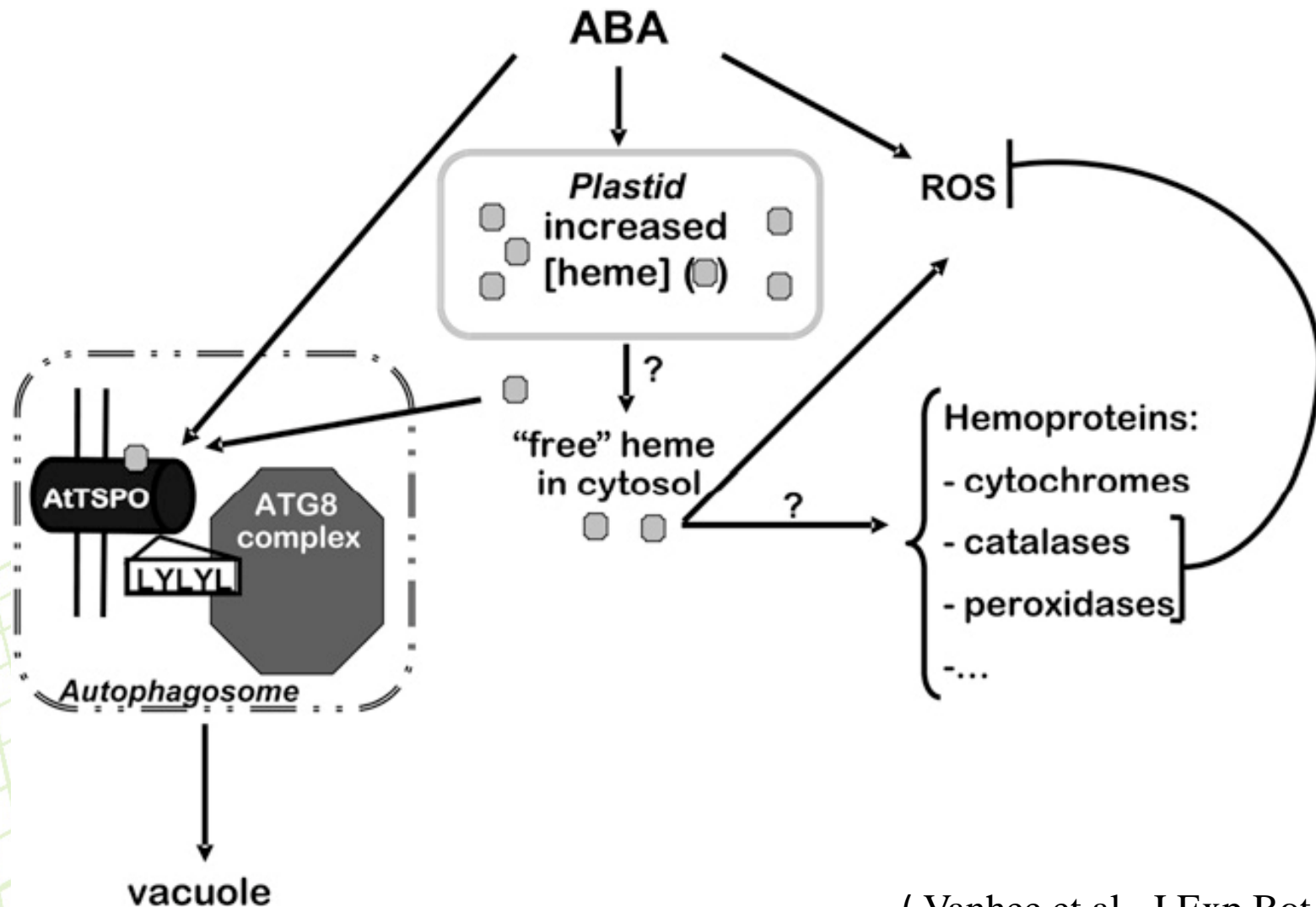


(Frank et al., PlantsJ,2007)



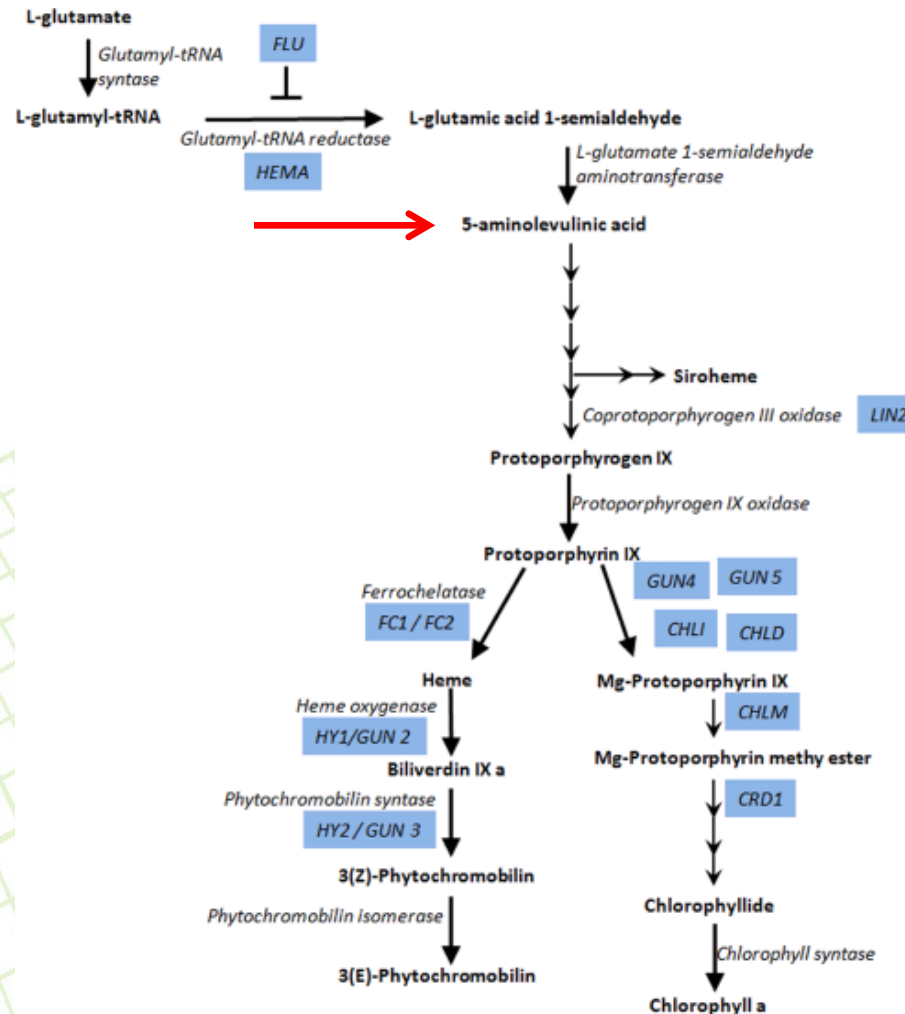
(Guillaumot et al., Plant J, 2009)

TSP0-Heme 通路研究进展



(Vanhee et al., J Exp Bot, 2011)

卟啉代谢通路



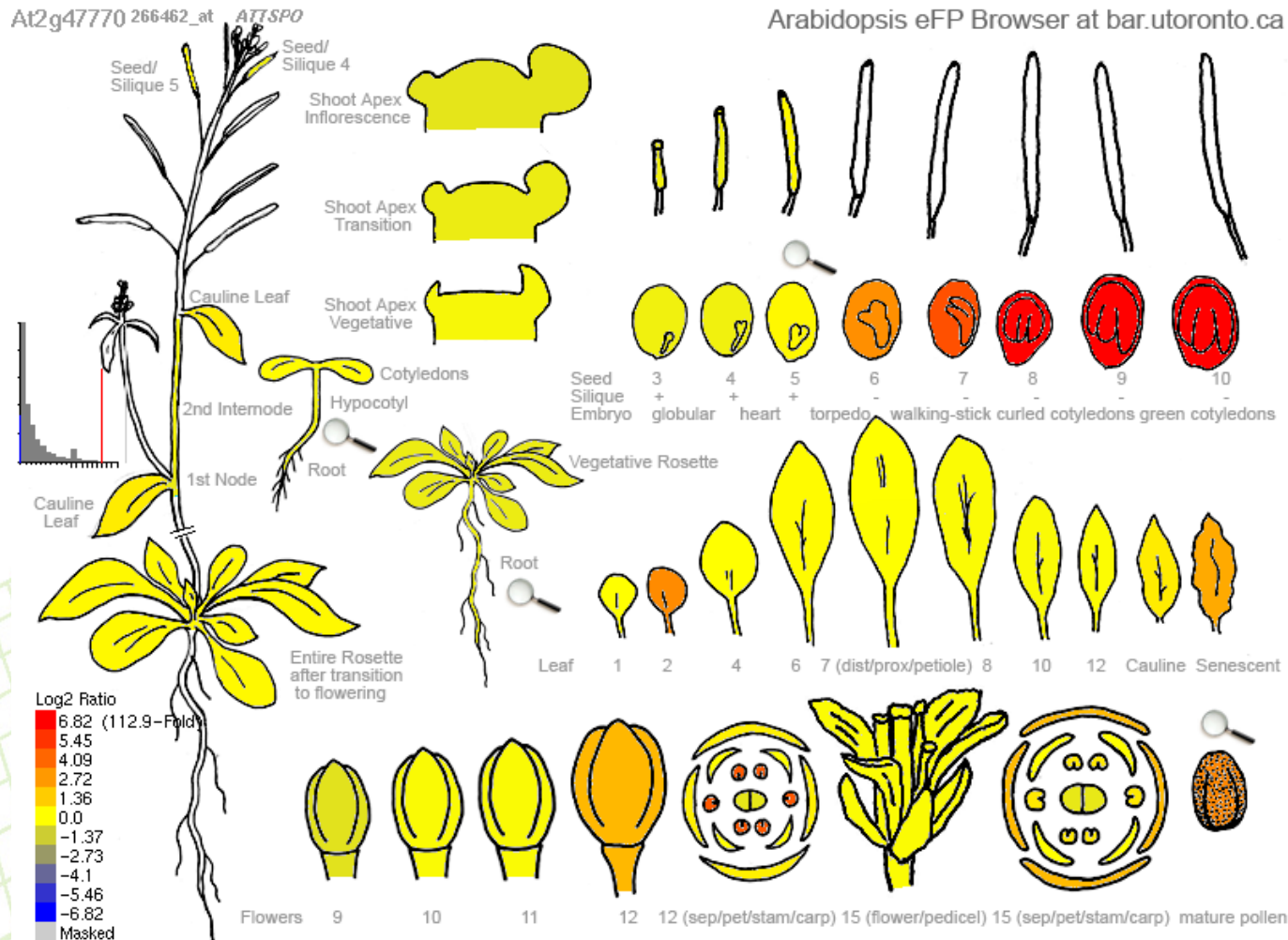
(Balsemão-Pires et al. BMC Plant Biology ,2011)



生物信息学分析



拟南芥中TSPO组织表达特异性

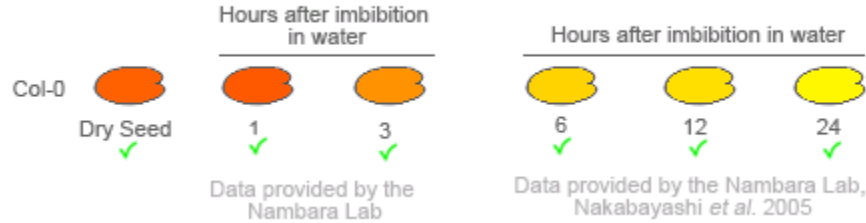


eFP Browser by B. Vinegar, drawn by J. Alls and N. Provart. Data from Gene Expression Map of Arabidopsis Development: Schmid et al., 2005, Nat. Gen. 37:501
Note that data displayed in this series were normalized by the RMA method and as such the levels are not directly comparable with other eFP Browser views.

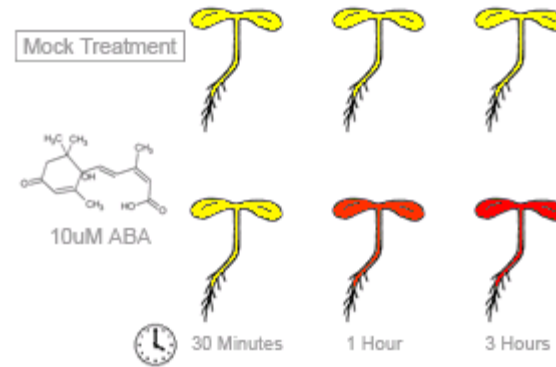
(Arabidopsis eFP Browser, <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>)

拟南芥中TSP0胁迫表达模式

Germination These seeds were after-ripened 2-4 months, and not stratified



ABA



Osmotic (300 mM Mannitol)



Salt (150mM NaCl)



(Arabidopsis eFP Browser, <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>)

拟南芥中顺式作用元件研究

Chr2:19568464..19569507
←-----→
19568.5k 19568.6k 19568.7k 19568.8k 19568.9k 19569k 19569.1k 19569.2k 19569.3k 19569.4k 19569.5k

Protein Coding Gene Models
AT2G47770.1 (TSPD,ATTSPD)

```
>tspo promoter 464nt
+ TGCACATTC GAGGGATACG CAAGCAAAAA AAAAAAGGAG GTAAATAGCT CTCACATGGA ACATAGGCGA
- ACGTGTAAG CTCCTATGC GTTCGTTTTT TTTTTTCCTC CATTATCGA GAGTGTACCT TGTATCCGCT
+ TAATAAGTGC GACTAGGATT GCATCAGCAA GAATTAGCGC GAATGCGAAT GGAAGTGCAG GTTTTTTGAA
- ATTATTCACG CTGATCCTAA CGTAGTCGTT CTTAATCGCG CTTACGCTTA CCTTGACGTC CAAAAA ACTT
+ TAGATCGGAT CGATTCGTCT CCTTCCCCAG CCGACGGCTA CGAGAAGCTC TCAAAC TCGC CGGTGATGAG
- ATCTAGCCTA GCTAAGCAGA GGAAGGGGTC GGCTGCCGAT GCTCTTCGAG AGTTTGAGCG GCCACTACTC
+ GCGCCCGCCA TGAAAACAGA GCAAATCGCA TCAGCGTCTA GCCAACGCCG CGTAACAGAC AACTACTTCC
- CGCGGGCGGT ACTTTTGTCT CGTTTAGCGT AGTCGCAGAT CGGTTGCGGC GCATTGTCTG TTGATGAAGG
+ ATATTACTAC TCTTCTAATT AGCCCAAATT AAATGAGCCT ATTGGGCTTC TTGTCTTAGT CGGTGTAGAG
- TATAATGATG AGAAGATTAA TCGGGTTTAA TTTACTCGGA TAACCCGAAG AACAGAATCA GCCACATCTC
+ CCCAATTGTT GTTTTATTTT TTAATAATGC AAAAGTATTA AGCGATAAAT AAATAAGCAT CGCAATCGTC
- GGGTTAACAA CAAAATAAAA AATTATTACG TTTTCATAAT TCGCTATTTA TTTATTCGTA GCGTTAGCAG
+ CAAAAC TGT GTGTATGCAT CAGACATGAG CATATAGAGT AAG
- GGT TTTGACA CACATACGTA GTCTGTACTC GTATATCTCA TTC
```

Factor or Site Name	Loc. (Str.)	Signal Sequence	SITE #
MYBST1	15 (+)	GGATA	S000180
GATABOX	16 (+)	GATA	S000039
MARTBOX	27 (-)	TTWTWTTWT	S000067
MARTBOX	28 (-)	TTWTWTTWT	S000067
PYRIMIDINEBOXOSRAMY1A	34 (-)	CCTTT	S000259
DOFCOREZM	35 (+)	AAAG	S000265
GT1CONSENSUS	41 (+)	GRWAAW	S000198
GTGANTG10	53 (-)	G TGA	S000378
EBOXBNNAPA	54 (-)	CANNTG	S000144
MYCOSENSUSAT	54 (-)	CANNTG	S000407
MYCATERD1	54 (-)	CATGTG	S000413
EBOXBNNAPA	54 (+)	CANNTG	S000144
MYCATRD22	54 (+)	CACATG	S000174
MYCOSENSUSAT	54 (+)	CANNTG	S000407
PRE CONSCRHSP70A	68 (+)	SOGAYNRNNNNNNNNNNNNNNNNHHD	S000506
GATABOX	70 (+)	GATA	S000039
GT1CONSENSUS	70 (+)	GRWAAW	S000198
IBOXCORE	70 (+)	GATAA	S000199
CACFTTPPCA1	77 (-)	YACT	S000449
ARR1AT	87 (+)	NGATT	S000454
CAATBOX1	89 (-)	CAAT	S000028
PE2FNTRNR1A	108 (-)	ATTGCGC	S000455
E2FCOSENSUS	108 (-)	WTTSSCSS	S000476
INTRONLOWER	127 (+)	TGCAGG	S000086
ARR1AT	152 (+)	NGATT	S000454
SURECOREATSULTR11	158 (-)	GAGAC	S000499
DRECRTCOREAT	171 (+)	ROCGAC	S000418
CBFHV	171 (+)	RYCGAC	S000497
LTRCOREATCOR15	172 (+)	COGAC	S000153
PRE CONSCRHSP70A	172 (+)	SOGAYNRNNNNNNNNNNNNNNNNHHD	S000506
HEXAMERATHA	173 (-)	COGTCG	S000146
CGAOGOSAMY3	173 (+)	CGAOG	S000205
ANAERO3CONSENSUS	204 (-)	TCATCAC	S000479
GTGANTG10	204 (+)	G TGA	S000378
EBCCRCAH1	232 (-)	GANTTNC	S000494

Factor or Site Name	Loc. (Str.)	Signal Sequence	SITE #
MYBST1	15 (+)	GGATA	S000180
GATABOX	16 (+)	GATA	S000039
MARTBOX	27 (-)	TTWTWTTWT	S000067
MARTBOX	28 (-)	TTWTWTTWT	S000067
PYRIMIDINEBOXOSRAMY1A	34 (-)	CCTTT	S000259
DOFCOREZM	35 (+)	AAAG	S000265
GT1CONSENSUS	41 (+)	GRWAAW	S000198
GTGANTG10	53 (-)	G TGA	S000378
EBOXBNNAPA	54 (-)	CANNTG	S000144
MYCOSENSUSAT	54 (-)	CANNTG	S000407
MYCATERD1	54 (-)	CATGTG	S000413
EBOXBNNAPA	54 (+)	CANNTG	S000144
MYCATRD22	54 (+)	CACATG	S000174
MYCOSENSUSAT	54 (+)	CANNTG	S000407
PRE CONSCRHSP70A	68 (+)	SOGAYNRNNNNNNNNNNNNNNNNHHD	S000506
GATABOX	70 (+)	GATA	S000039
GT1CONSENSUS	70 (+)	GRWAAW	S000198
IBOXCORE	70 (+)	GATAA	S000199
CACFTTPPCA1	77 (-)	YACT	S000449
ARR1AT	87 (+)	NGATT	S000454
CAATBOX1	89 (-)	CAAT	S000028
PE2FNTRNR1A	108 (-)	ATTGCGC	S000455
E2FCOSENSUS	108 (-)	WTTSSCSS	S000476
INTRONLOWER	127 (+)	TGCAGG	S000086
ARR1AT	152 (+)	NGATT	S000454
SURECOREATSULTR11	158 (-)	GAGAC	S000499
DRE CRTCOREAT	171 (+)	ROCGAC	S000418
CBFHV	171 (+)	RYCGAC	S000497
LTRCOREATCOR15	172 (+)	COGAC	S000153
PRE CONSCRHSP70A	172 (+)	SOGAYNRNNNNNNNNNNNNNNNNHHD	S000506
HEXAMERATHA	173 (-)	COGTCG	S000146
CGAOGOSAMY3	173 (+)	CGAOG	S000205
ANAERO3CONSENSUS	204 (-)	TCATCAC	S000479
GTGANTG10	204 (+)	G TGA	S000378
EBCCRCAH1	232 (-)	GANTTNC	S000494

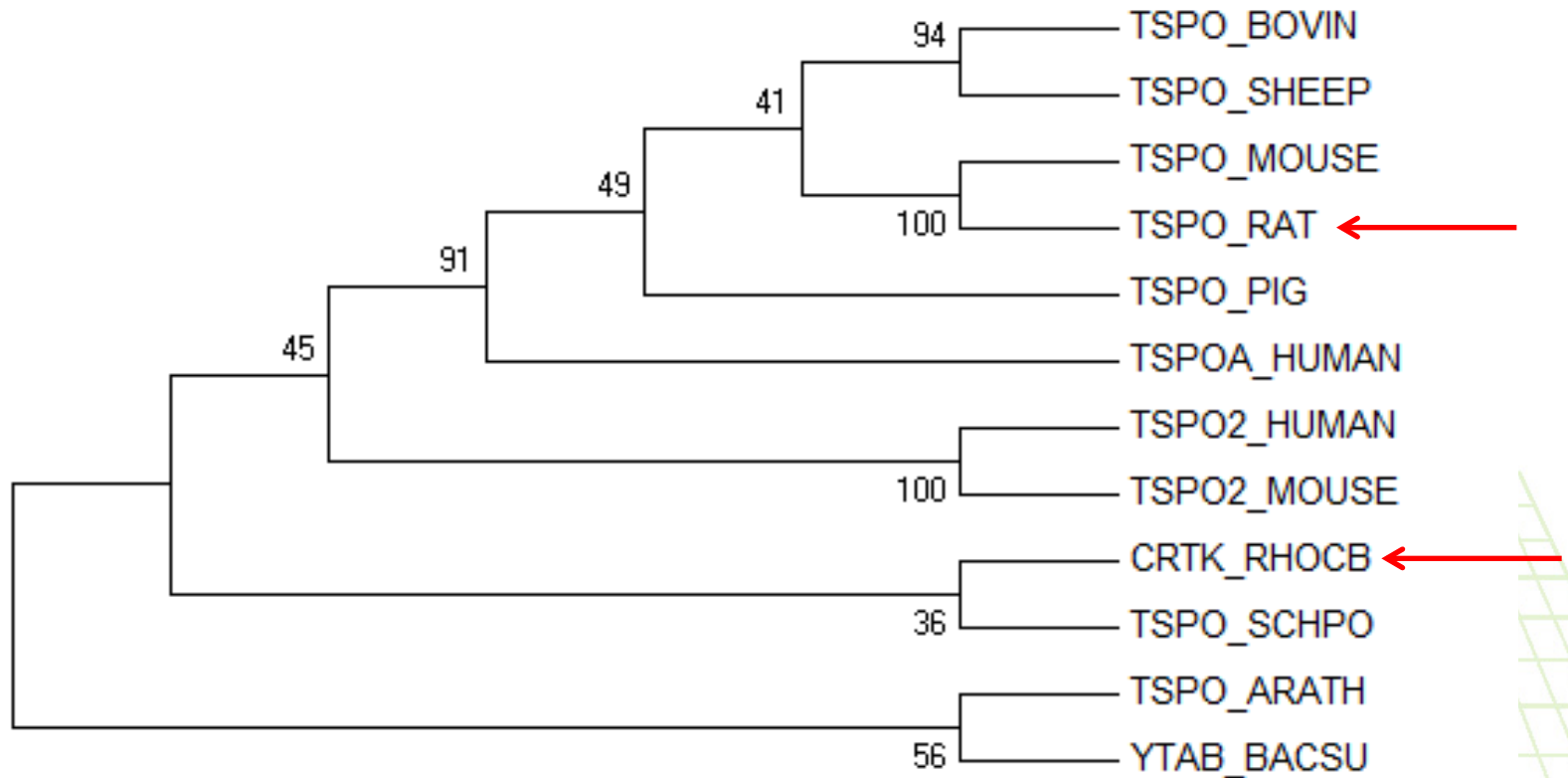
9个PRE元件, 6个DRE元件

(New PLACE, <https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?sid=&pj=0&lang=en&action=page&page=analysis>)

物种间TSPO 的多序列比对

Species/Abbrv	Group Name	*
1. sp O82245 TSPO_ARATH		MDSQDIRYRGGDDRDAAITAMAETEKSAADDNKGKRDQKRAMAKRGLKS
2. sp P17057 CRTK_RHOCE		M-----SLILFA
3. sp Q5TGU0 TSPO2_HUMAN		M-----RLQG
4. sp Q9CRZ8 TSPO2_MOUSE		M-----QLQG
5. sp P30536 TSPOA_HUMAN		M-----APRWVEA
6. sp P30535 TSPO_BOVIN		M-----APRWVEA
7. sp P50637 TSPO_MOUSE		M-----ESWVEA
8. sp P16257 TSPO_RAT		M-----SSWVEA
9. sp Q6UN27 TSPO_PIG		M-----APRWLEA
10. sp O94327 TSPO_SCHPO		M-----DLNYQVFSISK
11. sp Q9GMC9 TSPO_SHEEP		M-----APRWVEA
12. sp O34694 YTAB_BACSU		M-----KKIVG

TSPO 系统发生树分析



蛋白质序列分析

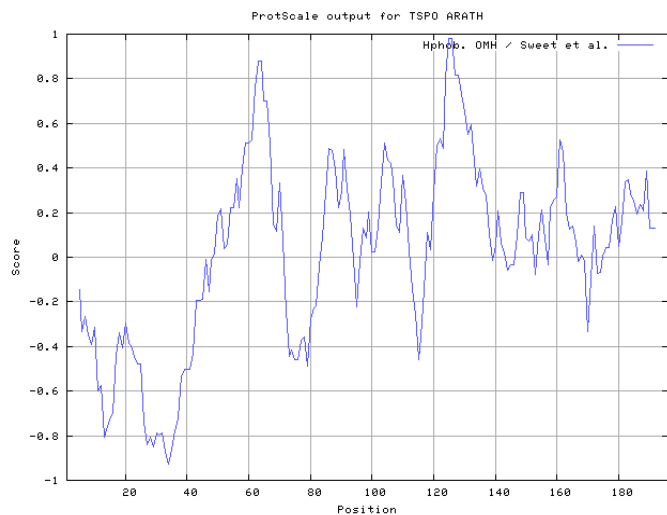
Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	73	37.245
Small	(A+B+C+D+G+N+P+S+T+V)	114	58.163
Aliphatic	(A+I+L+V)	74	37.755
Aromatic	(F+H+W+Y)	23	11.735
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	127	64.796
Polar	(D+E+H+K+N+Q+R+S+T+Z)	69	35.204
Charged	(B+D+E+H+K+R+Z)	37	18.878
Basic	(H+K+R)	23	11.735
Acidic	(B+D+E+Z)	14	7.143

有重复序列，但是在ExpASy
中的motif搜索工具没有找到
结果

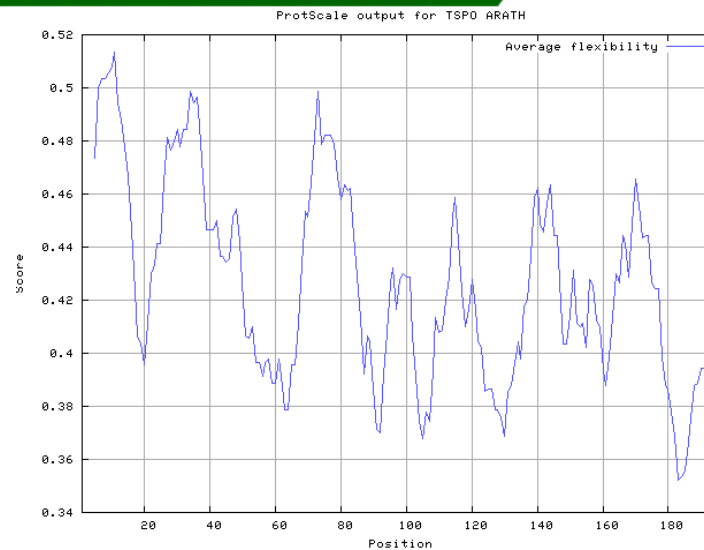
PEPSTATS, http://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/

Dotlet, <http://myhits.isb-sib.ch/cgi-bin/dotlet>

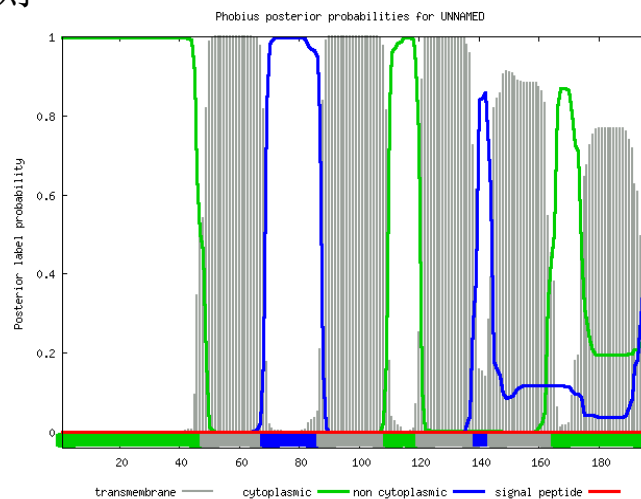
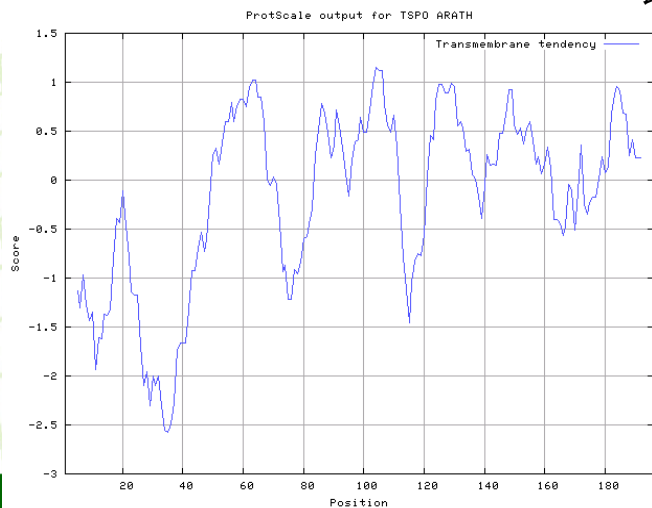
疏水性分析



柔性分析



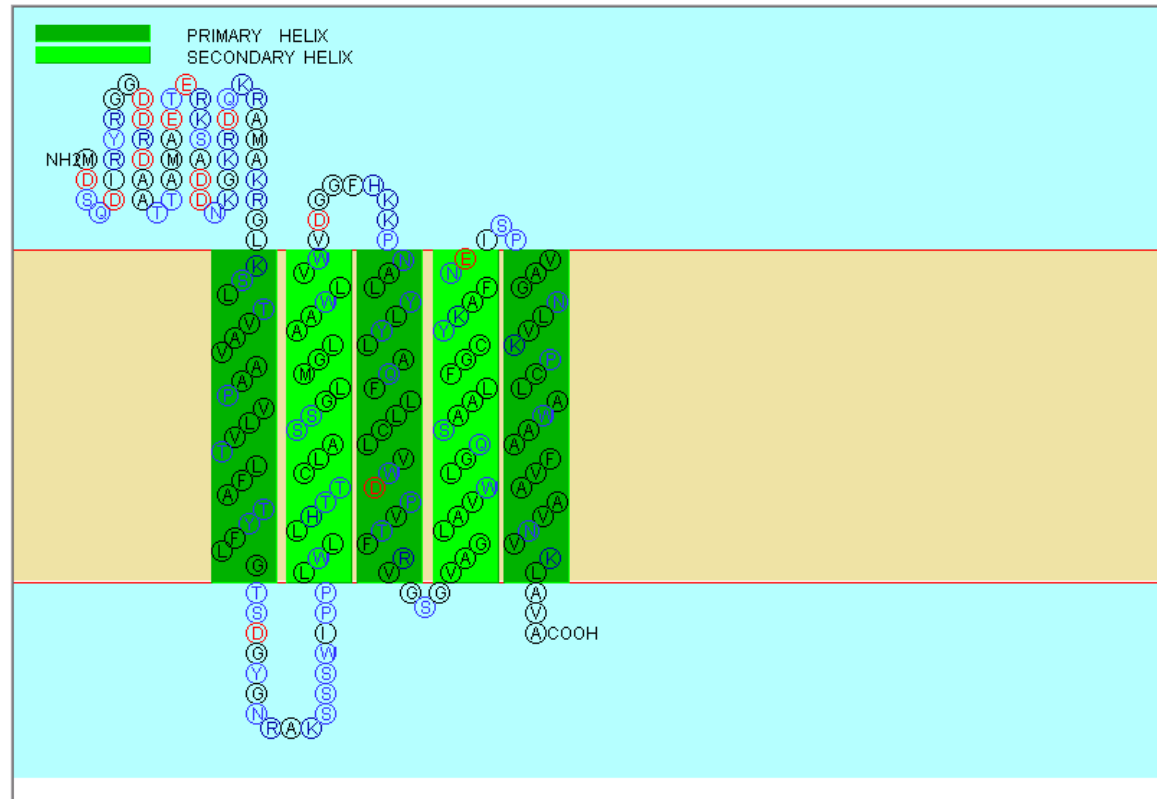
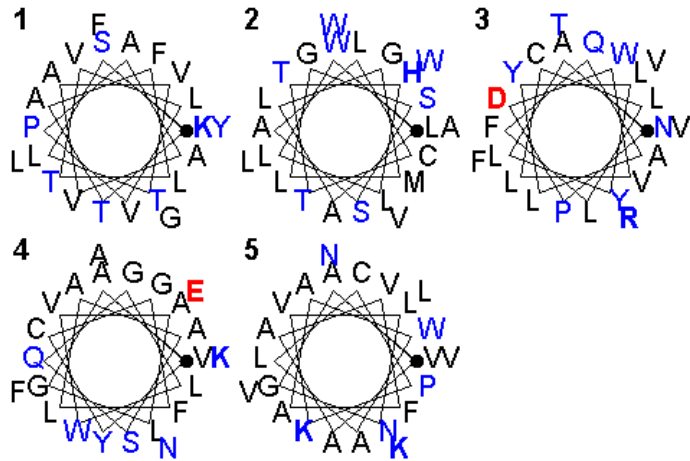
跨膜结构域预测



跨膜区预测

[Helical wheel diagram of predicted segments]

Hydrophobic residue: Black
 Polar residue: Blue
 Charged residue: Bold blue(+) Bold red(-)



(Sosui, http://harrier.nagahama-i-bio.ac.jp/sosui/cgi-bin/adv_sosui.cgi)

TSPO 定位预测分析

```
### targetp vl.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions not included.  
Using PLANT networks.
```

Name	Len	cTP	mTP	SP	other	Loc	RC
Sequence	196	0.077	0.109	0.116	0.862	_	2
cutoff		0.000	0.000	0.000	0.000		

```
### targetp vl.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions not included.  
Using PLANT networks.
```

Name	Len	cTP	mTP	SP	other	Loc	RC
Sequence	176	0.166	0.116	0.018	0.958	_	2
cutoff		0.000	0.000	0.000	0.000		

```
### targetp vl.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions not included.  
Using PLANT networks.
```

Name	Len	cTP	mTP	SP	other	Loc	RC
Sequence	155	0.002	0.069	0.917	0.107	S	1
cutoff		0.000	0.000	0.000	0.000		

```

5724 multiple located sequences are accepted
ProtComp Version 9.0. Identifying sub-cellular location (Plant)
Seq name: test sequence, Length=196
Significant similarity by DBSCAN-P - NONE
Predicted by Neural Nets - Extracellular (Secreted) with score 0.9
Integral Prediction of protein location: Extracellular (Secreted) with score 2.4
Location weights:      LocDB / PotLocDB / Neural Nets / Pentamers / Integral
Nuclear                0.0 / 0.0 / 0.00 / 0.00 / 0.00
Plasma membrane       0.0 / 0.0 / 0.91 / 0.35 / 0.00
Extracellular          0.0 / 0.0 / 0.91 / 2.66 / 2.37
Cytoplasmic            0.0 / 0.0 / 0.00 / 0.20 / 1.42
Mitochondrial         0.0 / 0.0 / 0.00 / 0.72 / 1.36
Endoplasm. retic.     0.0 / 0.0 / 0.00 / 1.95 / 0.00
Peroxisomal           0.0 / 0.0 / 0.91 / 0.00 / 2.22
Golgi                  0.0 / 0.0 / 0.27 / 0.33 / 0.84
Chloroplast            0.0 / 0.0 / 0.00 / 0.00 / 1.79
Vacuolar               0.0 / 0.0 / 0.00 / 0.00 / 0.00

```

```

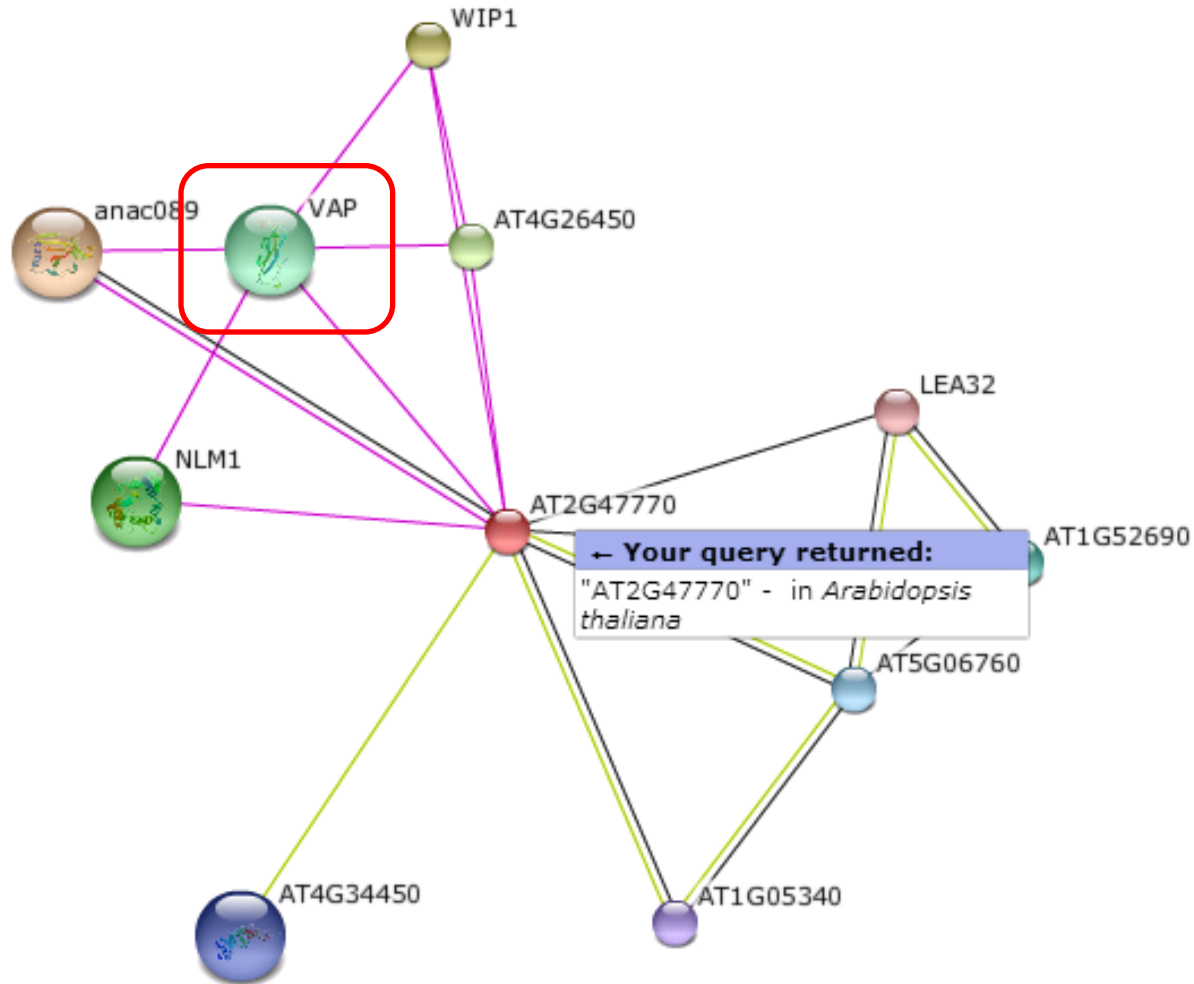
The protein is possibly multilocated: Chloroplast_and_Golgi_and_Endoplasmic_Reticulum due to SBLAST search in MultiLocDB
*****

```

ProtComp – Version

(<http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=proloc>)

TSPO 互作蛋白分析

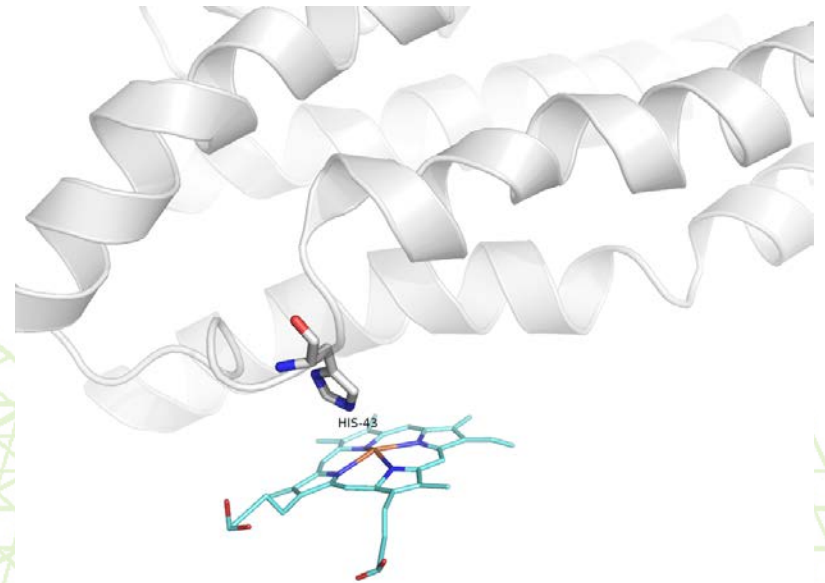
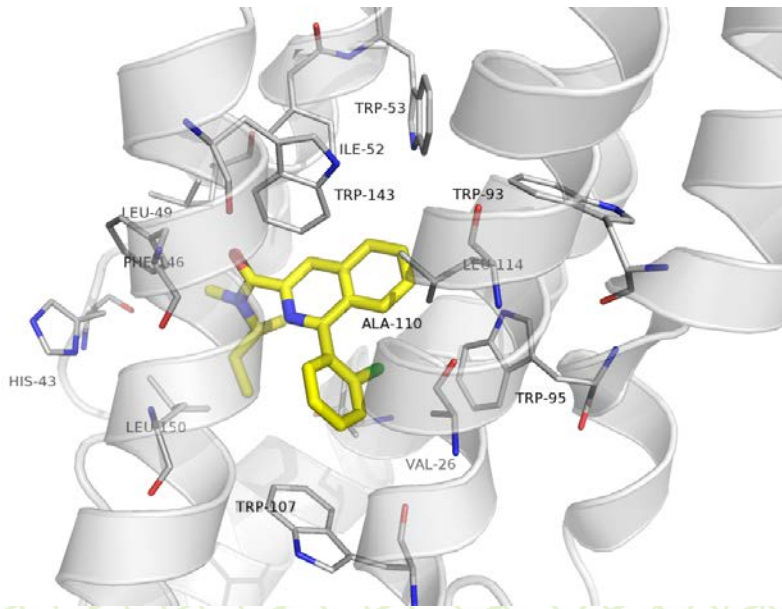


STRING 9.1

http://stringdb.org/newstring.cgi/show_network_section.pl?taskId=cu4pv_GyuTzV&interactive=yes&advanced_menu=_unassigned&network_flavor=evidence

结构预测

- Structure of Mouse Tspo

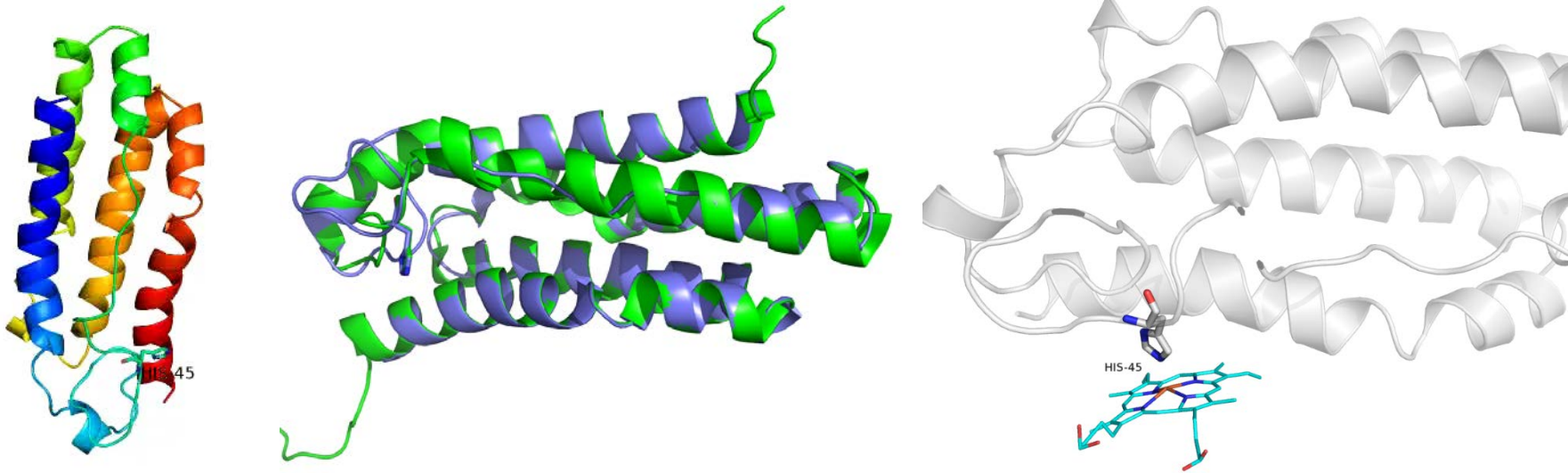


左：结合高亲和力配基的小鼠Tspo，内部空腔：(1)多为疏水性残基；(2)较窄，无法装载血红素；

右：预测小鼠Tspo血红素结合位点，His43。

结构预测

- Predicted Structure of AtTspo




左：利用Swiss-Model同源模建所得拟南芥Tspo预测结构；
中：拟南芥Tspo预测结构和模板比对图（绿色：小鼠Tspo；蓝色：预测拟南芥Tspo）；
右：预测拟南芥Tspo血红素结合位点，His45.

总结

- TSPO能被盐，渗透胁迫和ABA处理诱导
- TSPO可能在ABA信号转导通路和卟啉代谢途径行使特定功能, TSPO参与ROS的清除
- N端多出的40多个氨基酸可能与定位相关
- TSPO为膜定位蛋白

致谢

- 感谢罗老师的指导
- 感谢导师范六民教授的指导
- 感谢组员的合作参与
- 感谢PKU14F全体同学的陪伴



谢谢



The image is a title slide for a Q&A session. It features a white background with a light green grid pattern that recedes into the distance. The text 'Q&A' is centered in a bold, blue, sans-serif font with a slight drop shadow. The top and bottom of the slide are framed by green borders; the top border is a solid light green, and the bottom border is a solid dark green. In the top right corner, there is a circular inset showing a photograph of lush green trees.

Q&A