

白菜叶片卷曲候选基因KAN1的生物信息学分析

*The Bioinformatics Analysis of Candidate Gene BrKAN1
Related to Leaf Curvature in Brassica rapa*



小组：G12

组员：伦尧尧 付丽霞

魏明明 单天雷

报告人：付丽霞



内容简介

- 1、背景介绍
- 2、KAN1的核酸序列分析
- 3、KAN1的氨基酸序列分析
- 4、KAN1的蛋白质结构分析
- 5、总结与计划



背景介绍

白菜属于十字花科芸薹属作物，起源于中国，是我国栽培面积最大的蔬菜，栽培历史悠久，分布广阔，经过长期的进化和人工选择，遗传变异十分丰富，形成了许多亚种、变种及生态型。从形态学上可以将白菜分为两大类：结球白菜和非结球白菜。





散叶变种

1



半结球变种

2



花心变种

3



结球变种

4



➤ 结球白菜的生活周期



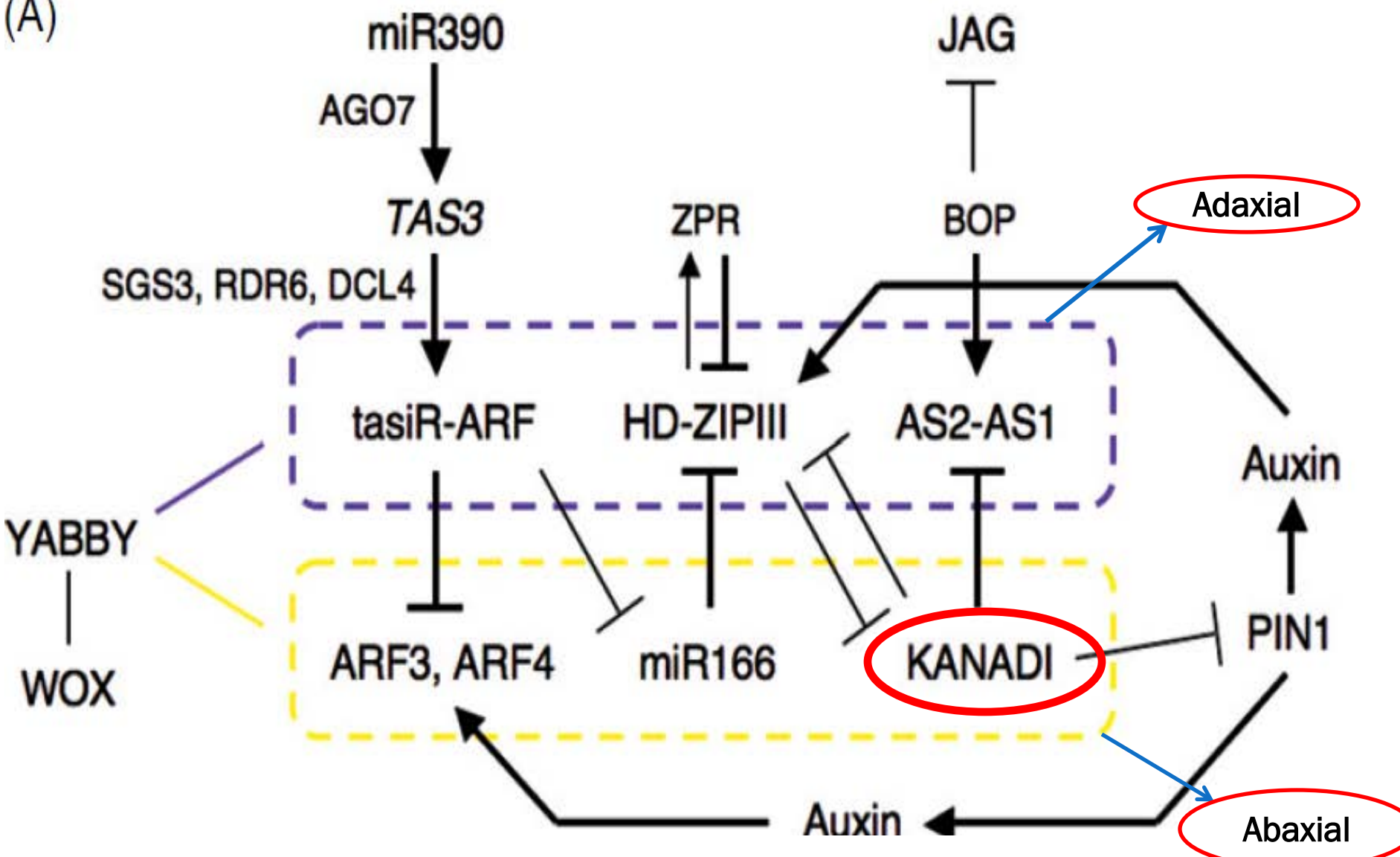
► 研究目的及意义

在结球白菜叶球形成过程中，叶片卷曲是最显著的特征之一，直接影响其产量和商品品质。因此，鉴定与白菜叶片卷曲相关的基因，揭示叶片卷曲形成的遗传规律及基因调控机制，不仅有助于阐明大白菜叶球形成的分子机理，而且对遗传改良结球白菜的商品品质及提高其产量方面具有重要的理论意义和实用价值。



参与植物叶片背腹性建立的基因调控网络^[1]

(A)

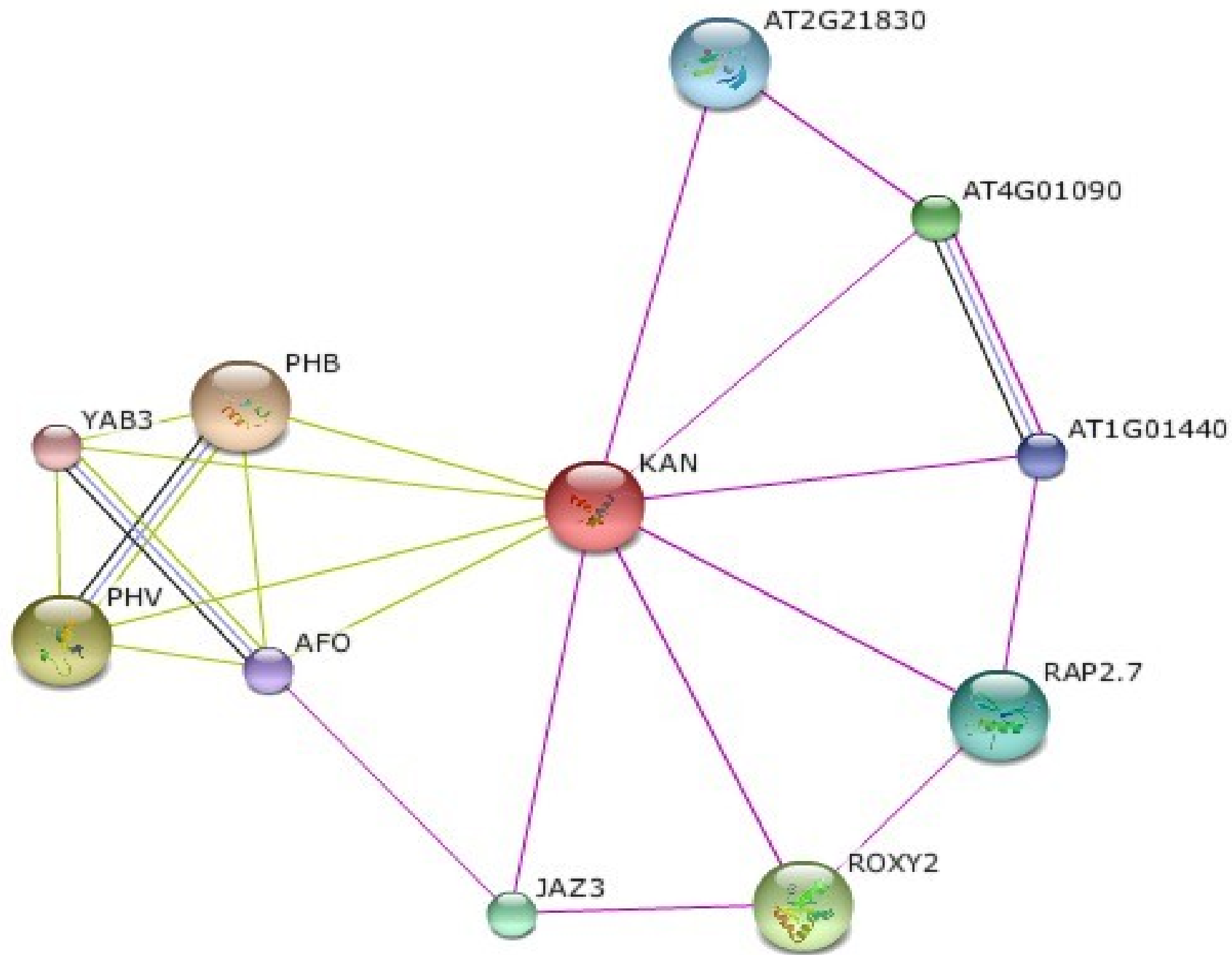


1、 KANADI是一类转录因子，编码 Myb-like GARP DNA 结合域，该家族中的KAN1、KAN2、KAN3、KAN4四个成员参与了叶片背侧的发育。（2-3）

2、在拟南芥中，KAN1是一类转录抑制因子，能够调节侧器官的极性发育。（1）它通过抑制侧器官背面细胞中调节腹面发育的基因AS2的作用来促进背侧的极性发育，其作用在叶片和心皮的发育中尤为重要；（2）能与KAN2在珠背极性分化过程中共同发挥作用。（3-5）

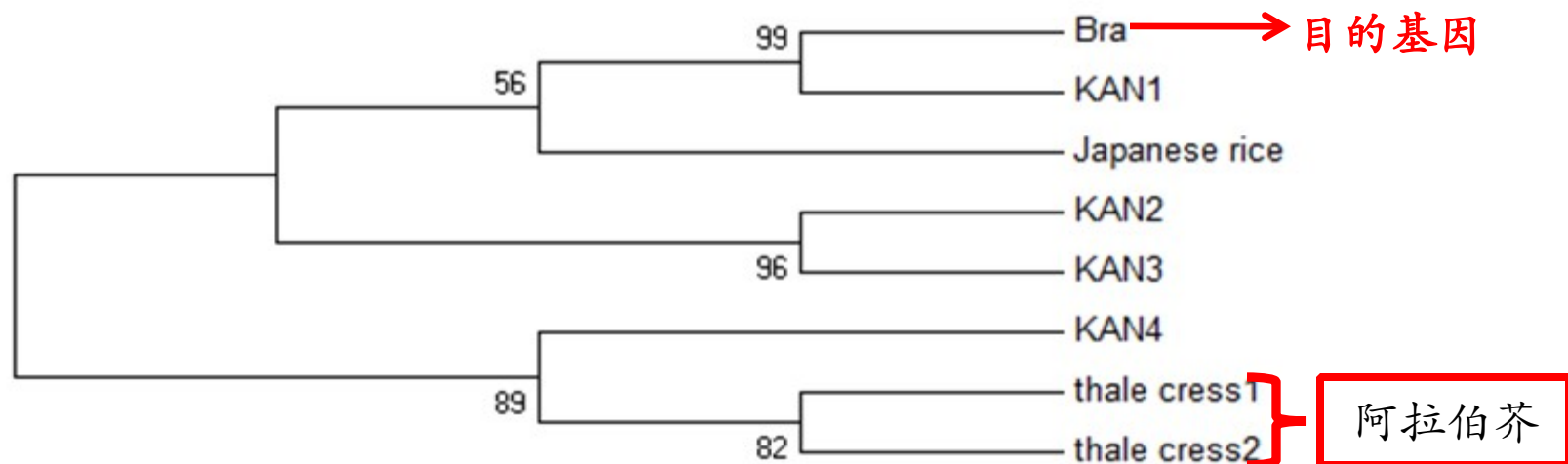


Entry	Entry name	Status	Protein names	Organism	Length
Q93WJ9	KAN1_ARATH	★	Transcription repressor KAN1	Arabidopsis thaliana (Mouse-ear cress)	403
Q9FJV5	KAN4_ARATH	★	Probable transcription factor KAN4	Arabidopsis thaliana (Mouse-ear cress)	276
Q9C616	KAN2_ARATH	★	Probable transcription factor KAN2	Arabidopsis thaliana (Mouse-ear cress)	388
Q94112	KAN3_ARATH	★	Probable transcription factor KAN3	Arabidopsis thaliana (Mouse-ear cress)	322



➤ 用目的蛋白的氨基酸序列在NCBI中进行比对并构建系统进化

	Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
<input checked="" type="checkbox"/>	RecName: Full=Transcription repressor KAN1; AltName: Full=Protein KANADI 1	375	375	93%	4e-127	79%	Q93WJ9.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Probable transcription factor KAN2; AltName: Full=Protein KANADI 2	157	157	29%	4e-43	84%	Q9C616.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Probable transcription factor RL9; AltName: Full=Protein ROLLED LEAF 9	158	158	28%	1e-42	88%	Q0J235.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Probable transcription factor KAN3; AltName: Full=Protein KANADI 3	135	135	63%	1e-35	44%	Q941I2.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Probable transcription factor KAN4; AltName: Full=Protein ABERRANT TESTA SHAPE; AltName: Full=Protein K	124	124	25%	6e-32	79%	Q9FJV5.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Putative Myb family transcription factor At1g14600	84.0	84.0	20%	8e-18	64%	Q700D9.2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	RecName: Full=Protein PHR1-LIKE 1; AltName: Full=Myb-like transcription factor 1	74.3	74.3	22%	7e-14	58%	Q8GUN5.1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>



白菜候选基因KAN1与拟南芥KAN1的cds序列比对:

Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
1235	3196.5	763/1235 (61.8%)	763/1235 (61.8%)	382/1235 (30.9%)

氨基酸序列比对:

Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
409	844.0	224/409 (54.8%)	242/409 (59.2%)	124/409 (30.3%)

```

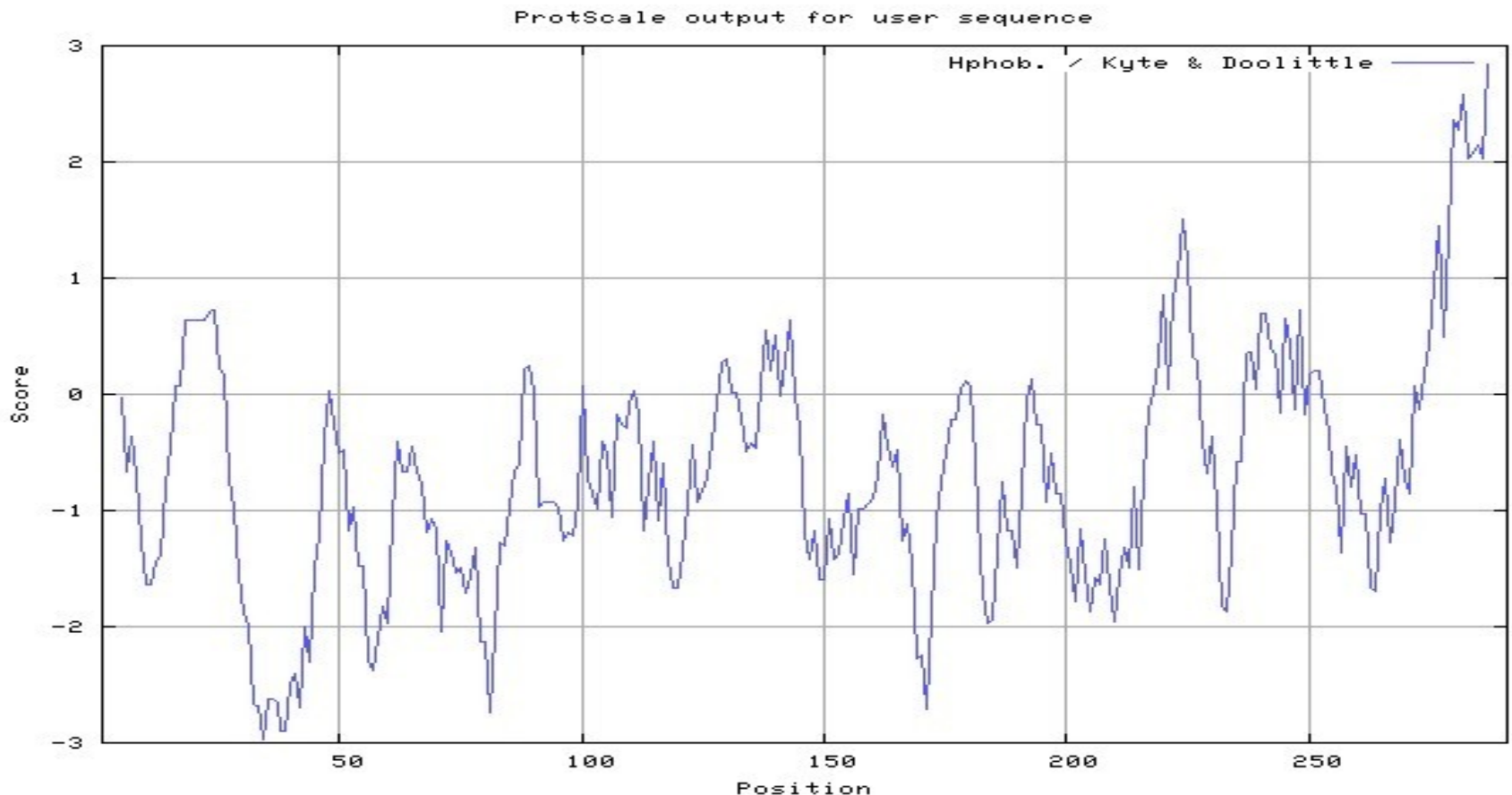
195  HNHHHHGMIRSRFLPKMPTKRSMRAPRMRWTSSLHARFVHAVELLGGHER      244
      |||
186  ---HHHGMIRSRFLPKMPTKRSMRAPRMRWRSSLHARFVHAVELLGGHER      232
      |||
245  ATPKSVLELMDVKDLTLAHVKSHLQMYRTVKTTNKPAASSDGSGEEMGI      294
      |||
233  ATPKSVLELMDVKDLTLAHVKSHLQMYRTVKTTNKPAASSVKRIEIVLLR      282
  
```

➤ 氨基酸序列分析

用ExPASy中的ProtParam分析:

M4CWL6_BRARP蛋白的分子量32839.2Da, 等电点为10.68, 不稳定系数为61.83。总平均疏水指数-0.633

➤ 用ProtScale分析疏水性



➤用PEPSTATS分析氨基酸组成:

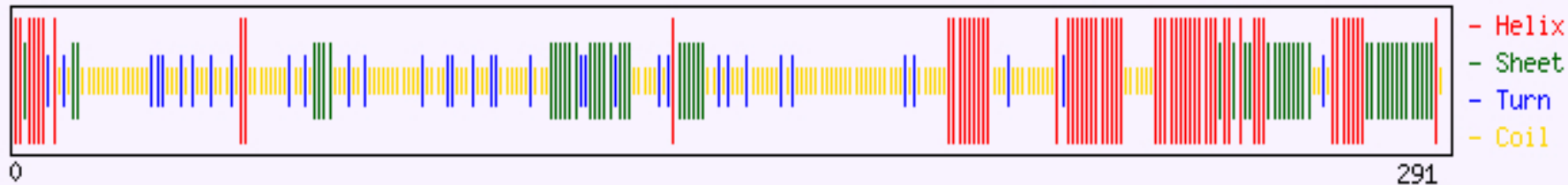
PEPSTATS of KAN1 from 1 to 291

Molecular weight = 32839.11 Residues = 291
Average Residue Weight = 112.849 Charge = 29.5
Isoelectric Point = 11.2578
A280 Molar Extinction Coefficient = 14650
A280 Extinction Coefficient 1mg/ml = 0.45
Improbability of expression in inclusion bodies = 0.946

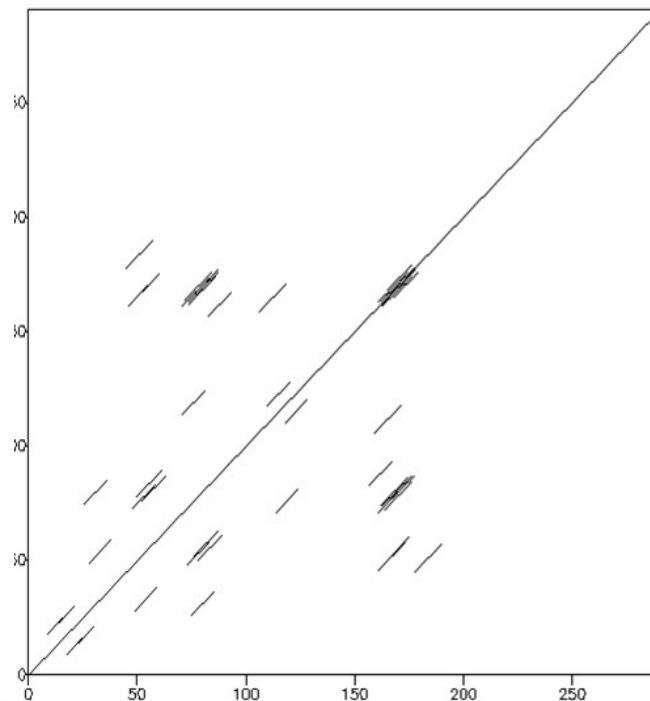
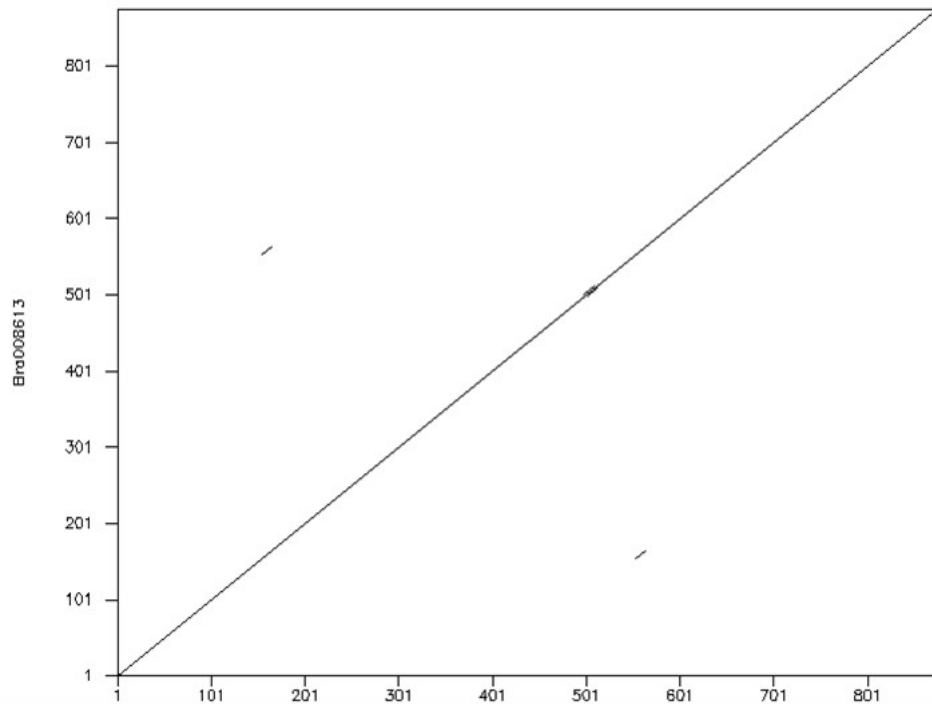
Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	77	26.460
Small	(A+B+C+D+G+N+P+S+T+V)	145	49.828
Aliphatic	(A+I+L+V)	70	24.055
Aromatic	(F+H+W+Y)	41	14.089
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	132	45.361
Polar	(D+E+H+K+N+Q+R+S+T+Z)	159	54.639
Charged	(B+D+E+H+K+R+Z)	78	26.804
Basic	(H+K+R)	60	20.619
Acidic	(B+D+E+Z)	18	6.186

➤ CFSSP对二级结构的预测及重复序列分析

(<http://www.biogem.org/tool/chou-fasman/>)



Dotup: fasta::699039:Bra008613 vs fasta::699040:Bra0086... Dotmatcher: fasta::699049:Bra008613.1 vs fasta::699050:B.
Sun 12 Jan 2014 21:45:07 (window size = 10, threshold = 23.00 12/01/14)

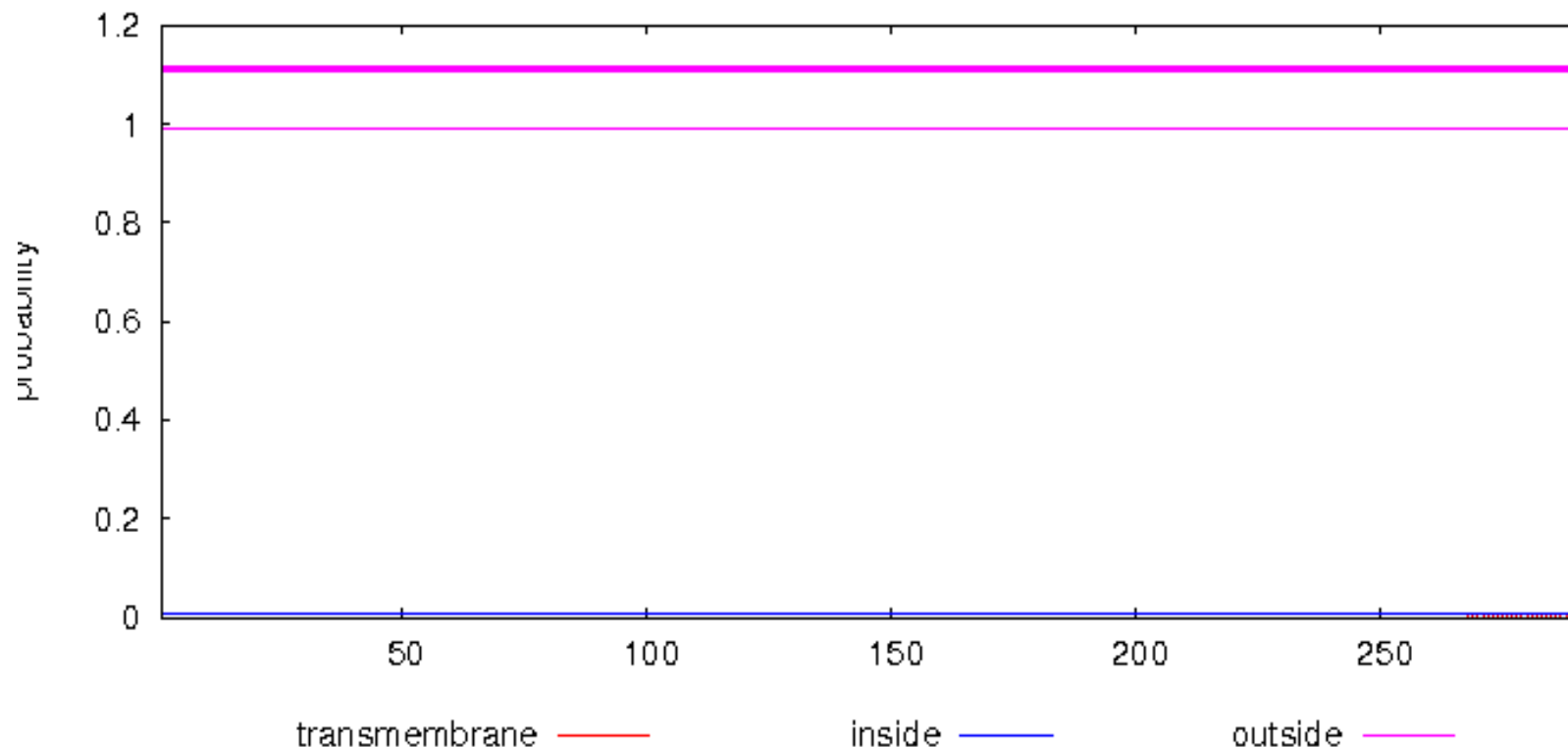


➤ TMHMM分析跨膜螺旋

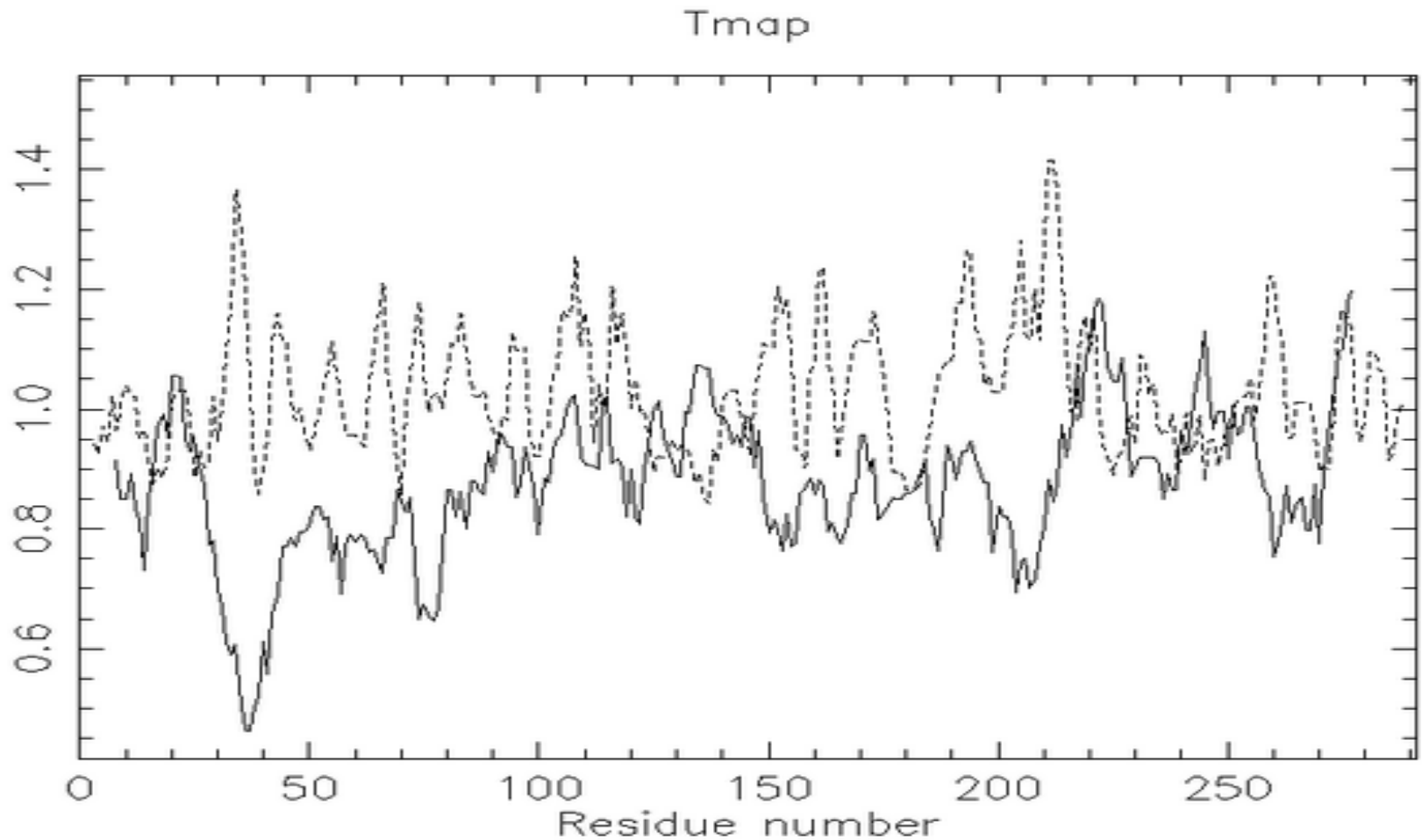
<http://www.cbs.dtu.dk/services/TMHMM/>

```
# WEBSEQUENCE Length: 291
# WEBSEQUENCE Number of predicted TMHs: 0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00779
# WEBSEQUENCE Exp number, first 60 AAs: 0
# WEBSEQUENCE Total prob of N-in: 0.00909
WEBSEQUENCE      TMHMM2.0      outside      1      291
```

TMHMM posterior probabilities for WEBSEQUENCE



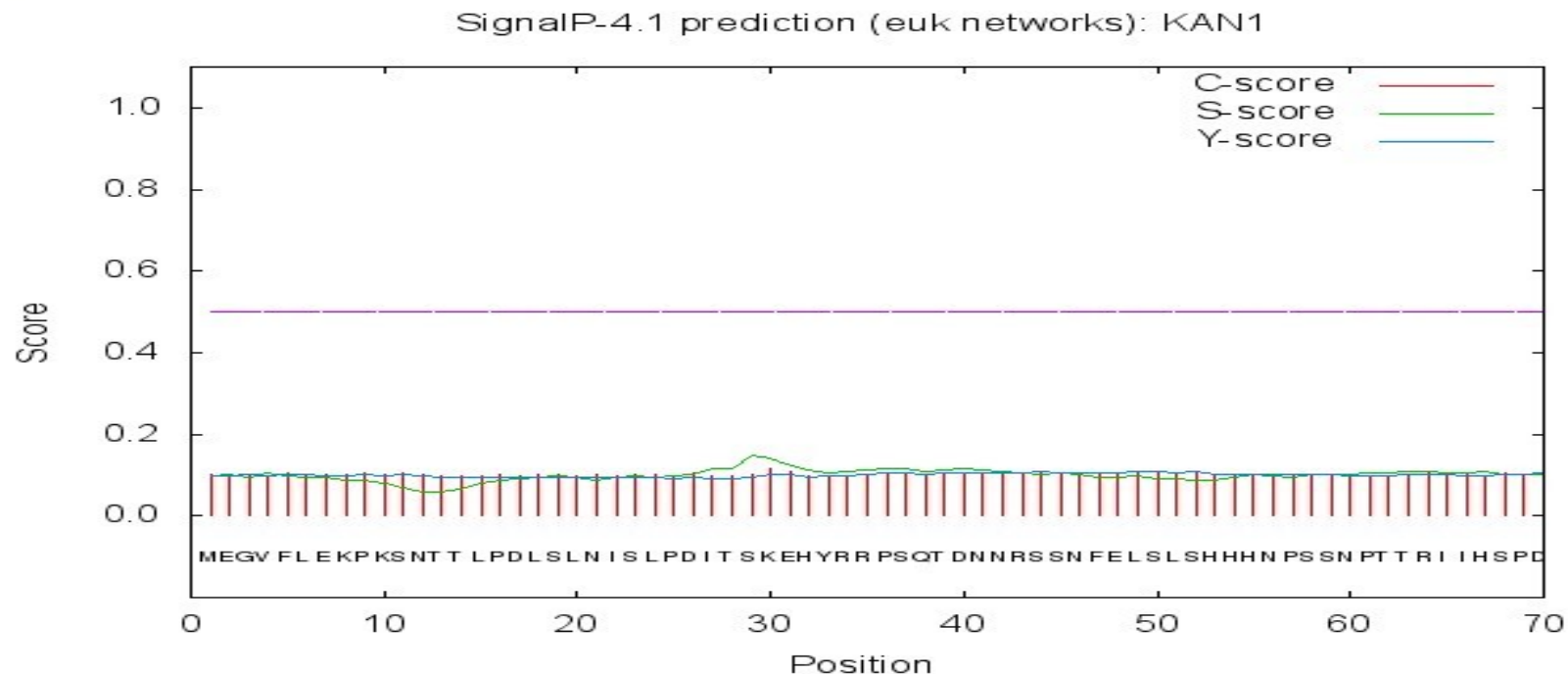
➤用Tmap进行跨膜螺旋预测



无跨膜螺旋区，排除是膜蛋白的可能性

➤ 信号肽预测：

```
# SignalP-4.1 euk predictions  
>KAN1
```



# Measure	Position	Value	Cutoff	signal peptide?
max. C	30	0.117		
max. Y	49	0.108		
max. S	29	0.148		
mean S	1-48	0.100		
D	1-48	0.104	0.450	NO

Name=KAN1 SP='NO' D=0.104 D-cutoff=0.450 Networks=SignalP-noTM

没有信号肽，不是分泌蛋白

➤ 亚细胞定位

<http://www.cbs.dtu.dk/services/TargetP/>)



TargetP 1.1 Server - prediction results

Technical University of Denmark

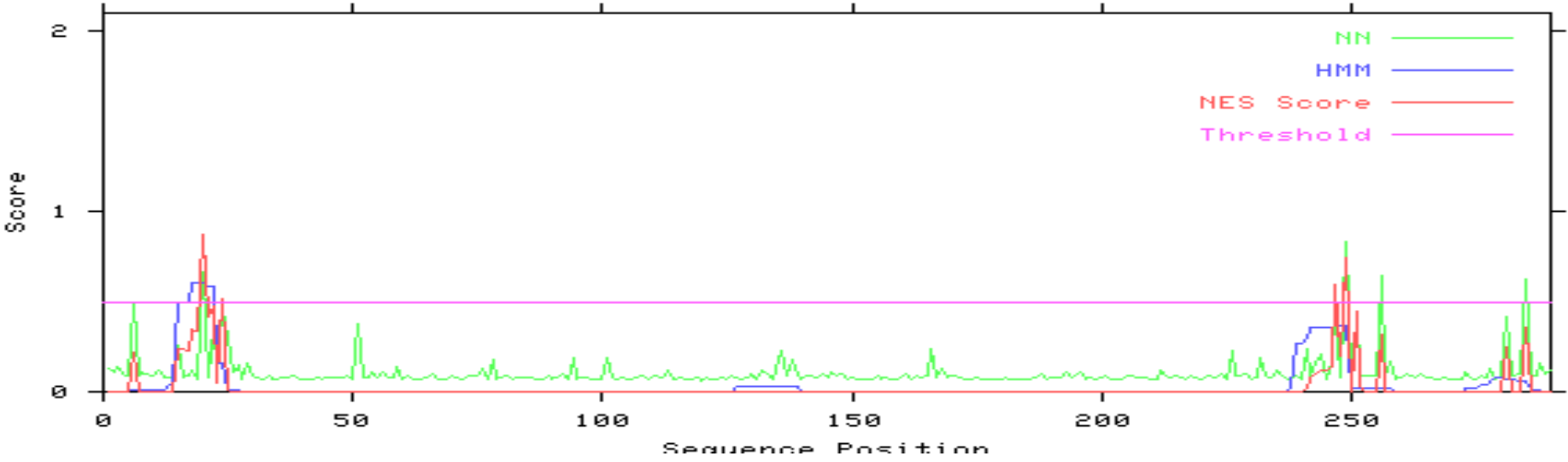
```
### targetp v1.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions not included.  
Using PLANT networks.
```

Name	Len	cTP	mTP	SP	other	Loc	RC
Bra008613.1	291	0.209	0.070	0.008	0.836	_	2
cutoff		0.000	0.000	0.000	0.000		

结果显示：该蛋白有可能位于细胞质基质中，排除位于叶绿体、线粒体的可能。

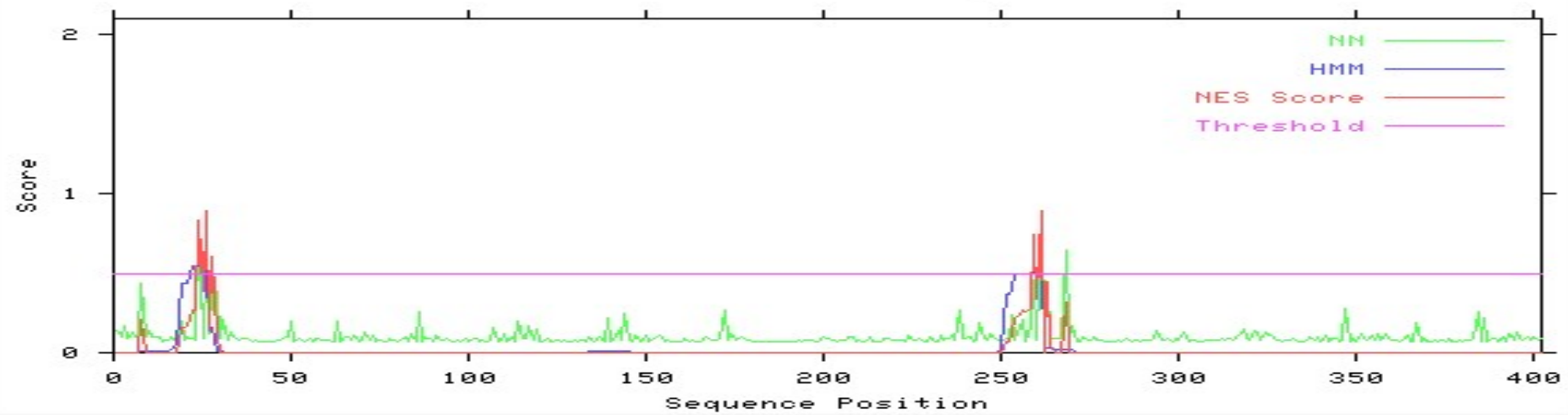
核定位信号预测

NetNES 1.1: Predicted NES signals in Bra008613.1



>KAN1 - NetNES 1.1 prediction

NetNES 1.1: Predicted NES signals in KAN1



20	24	247	249
L	L	L	L

24	26	28	259	261
L	I	L	L	L

➤ 用DNAMAN进行多序列比对

```

KAN1          ...MEGVFLEKPKSNTTLPDLSLNISLPDIT.....SKEHYR          34
XP_002533735.1 MSRNRKGGIFYGKQSLNQIPDLSLHISLPNSAPSS.....ICIGTNDADDDSSF          47
NP_001175589.1 .....MASATSTTPDLSLHISLPSGGAAVGTTAPPGLGNGGAVGGAGGGKAAGGG          50
gi|29837235    .....MASATSTTPDLSLHISLPSGGAAVGTTAPPGLGNGGAVGGAGGGKAAGGG          50
EPS72669.1     .....                                .....                                0
XP_004244407.1 .....                                .....                                0
XP_004136400.1 .....MELFPAQPDLSQLISPPNTN.....KPSSSW          26
Consensus

```

```

KAN1          RPSQTDNNRSSNFELSLSHHHNPSSNPTRIIHSP.....DPRNL.....          74
XP_002533735.1 HGWHKDDDDDDDDVDVGLKSHSDSSIKVGSNIQPVDDDIELSLANPTST.....          95
NP_001175589.1 DPWRRLDGSTASTELSLSTPQQQEGSTSAADVLPWRRLRQPTAAAAASVPVTLPTIPMDGS          110
gi|29837235    DPWRRLDGSTASTELSLSTPQQQEGSTSAADVLPWRRLRQPTAAAAASVPVTLPTIPMDGS          110
EPS72669.1     .....                                .....                                0
XP_004244407.1 .....                                .....                                0
XP_004136400.1 R.SRTSSSSSSQEDVDLEFWNNKSTPKHHLHHLHPH...SCFDLSLS.....          69
Consensus

```

```

KAN1          NLPHHQHYNNP.....ILNGGSLHQRVNGSDINSIHR.PIRGIPVYQNRSFPPFHQQTSLP          128
XP_002533735.1 ALEAESPWRRISNFASDGRRANCSKEDQARQRNLLSRANNGISVFEVSGFKSIKGFVY          155
NP_001175589.1 AAAARAPIRGVPVYSGGGGGGHPFLGGGGGDHRHNRLYNPYHSTAWPSSSLCSTSPAPAP          170
gi|29837235    AAAARAPIRGVPVYSGGGGGGHPFLGGGGGDHRHNRLYNPYHSTAWPSSSLCSTSPAPAP          170
EPS72669.1     .....                                .....                                0
XP_004244407.1 .....                                .....                                0
XP_004136400.1 NLTQINHHQYP.....NVSPSPTPNHHHHLLHHHHRYFVLDSPAHLFETLNKTHNRPSA          124
Consensus

```

```

KAN1          SLGGGGDLDQISILNSSSGYN..NAYRSLQSSPRLRGVPLHHHHHNHYGVVGS....TDS          182
XP_002533735.1 STGEIMDPRFCFNQMAYSSSCTPYPSTDNCSFPAFRIGTSYHHHHQYASGAGGGG..GGA          213
NP_001175589.1 PPPPAALDPTTSSLLSPSAYHRMLSSTGRLHGVLADTLRGYAGAAVAGSIGYGSAAAAA          230
gi|29837235    PPPPAALDPTTSSLLSPSAYHRMLSSTGRLHGVLADTLRGYAGAAVAGSIGYGSAAAAA          230
EPS72669.1     .....                                .....                                0
XP_004244407.1 .....                                .....                                0
XP_004136400.1 IFHHNTPPSSSLDPTPPSFPGQINGQDLGFLRPIRGIPVYHQTNNSVGTGNGGVVTPS          184
Consensus

```


KAN1	SSPHHHG.....MIRSRFLPKMP..TKRSMRAPRMRWRS...LHARFVHAVELLGGHERA	233
XP_002533735.1	AEVYGGG.....IIRSRFMPKLQ..NKRNMRAPRMRWTSS...LHARFVHAVELLGGHERA	264
NP_001175589.1	AAMGGYGGAGAGGGGFASSRFMPRLLPASRRSMRAPRMRWTSS...LHARFVHAVELLGGHERA	290
gi 29837235	AAMGGYGGAGAGGGGFASSRFMPRLLPASRRSMRAPRMRWTSS...LHARFVHAVELLGGHERA	290
EPS72669.1MRSRFLPKIP..AKRSMRAPRMRWTSS...LHARFVHAVELLGGHERA	43
XP_004244407.1MRSRFFSRFP..TKRSMRAPRMRWTSS...LHARFVHAVELLGGHERA	43
XP_004136400.1	AASHGQI.....MMRSRFLSRFP..AKRSMRAPRMRWTSS...LHARFVHAVELLGGHERA	235
Consensus	srf r mraprmrw s lharfvhav llgghera	

KAN1	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKTTNK...FAASSV..KRIEIVLLRVVLSG.	287
XP_002533735.1	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKSTDK...FAASS...DGSGDEDFLSVTTP.	317
NP_001175589.1	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKSTDK...FAAASGPMDGSGSGSGSGDDE..	345
gi 29837235	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKSTDK...FAAASGPMDGSGSGSGSGDDE..	345
EPS72669.1	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKTTDK...FAASSGHS DGSGEDDMSTVG...	97
XP_004244407.1	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKTTDKSAVGAASSG..QSEVFDNVSSGDNS.	100
XP_004136400.1	TPKSVLELMDVKDLTLAHVKSHLQMYRTVKTTDR...AAASSG..QSDLYENASSGDTSN	290
Consensus	tpksvlelmdvkdltlahvkshlqmyrtvk t aa s	

KAN1	...IIF... ..	291
XP_002533735.1ITQNSSHFLNPTRGSSVLENDDDNNVGYPPSTLWSNSSR.....	356
NP_001175589.1LLAGD...GRQATSSGADARRMSFTEHRSSSEGAASHAG.....GGGDGDC	389
gi 29837235LLAGD...GRQATSSGADARRMSFTEHRSSSEGAASHAG.....GGGDGDC	389
EPS72669.1SGGGLRQFIGSPPRDSANFQSSTPP.TLWSNSSSGRDGWLQMNGG.....	142
XP_004244407.1	.DDLLEIKN...SNKSELSVQQGRQNMYNQEKDYGGLWSNSSSRESWHMHGKLGDPGNL	156
XP_004136400.1	IEDIIFDIQKPTIGRQENSIQQTSRQNP HQDKDYHALWSNSSRN.....DVDPKS	341
Consensus		

KAN1	291
XP_002533735.1	356
NP_001175589.1	SSSAVNSDTIRARSNSPSVQSSPRLEI.....	416
gi 29837235	SSSAVNSDTIRARSNSPSTSIETWDCVWCCSLKEKS.....	425
EPS72669.1	142
XP_004244407.1	PSLEKNMEAKCTSYEGISTEVSSSSITQASPKKP NLEFTLGI PNI	201
XP_004136400.1	SNNERLSEISS.....STDIICG.VTNNNPLKPNLEFTLGRPI.	378
Consensus		

➤ 结构域预测-SMART



Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
low complexity	49	63	N/A
low complexity	130	137	N/A
low complexity	167	173	N/A

Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
HLH	204	261	1510	threshold

HLH Domain

HLH domain

This is a SMART **HLH** domain ([full annotation](#)).



Position: 204 to 261

E-value: 1509.32657597445 ([HMMER2](#))

SMART ACC: [SM000353](#)

Definition: helix loop helix domain

Description:

Interpro abstract (IPR011598): A number of eukaryotic proteins, which probably are sequence specific DNA-binding proteins that act as transcription factors, share a conserved domain of 40 to 50 amino acid residues. It has been proposed ...([full abstract](#))

GO function: protein dimerization activity ([GO:0046983](#))

HLH domain sequence (58 aa):

[Submit to BLAST](#)

[Align with the SMART alignment](#)

[Copy to clipboard](#)

SMRAPMRWRSSSLHARFVHAVELLGGHERATPKSVLEELMDVKDLTLAHVKSHLQMYRT

KAN1蛋白的204-261aa是**Myb DNA-binding**结构域:

APRMRWRSSSLHARFVHAVELLGGHERATPKSVLEELMDVKDLTLAHVKSHLQ
MY

➤ 用Swiss Model做蛋白质同源模建

Workunit: P000004 KAN1 - Overview



Print/Save this page as

Model Summary



Model information:

Modelled residue range: 204 to 264
Based on template: [1irzA] (99.9 Å)
Sequence Identity [%]: 44.26
Evaluate: 3.80e-26

Quality information: [details]▼

QMEAN Z-Score: -0.76

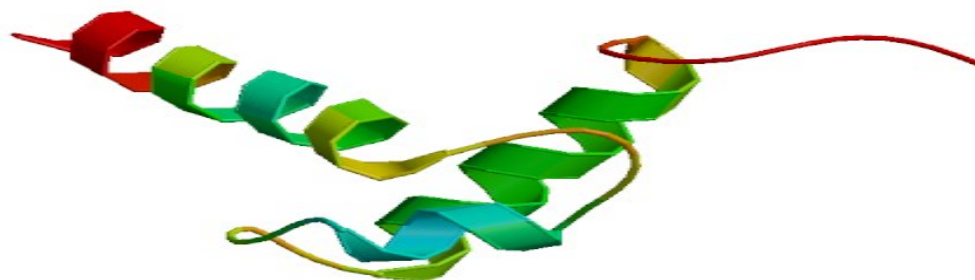


Quaternary structure information: [details]▼

Template (1irz): MONOMER
Model built: SINGLE CHAIN

Ligand information: [details]▼

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



➤ 结论：

- (1) 白菜叶片卷曲候选基因KAN1与拟南芥KAN1基因的相似性较高；
- (2) 二级结构分析可知，在204-261aa处有 α 螺旋区
- (3) 跨膜螺旋及信号肽分析可知，目的蛋白不是膜蛋白，不含有信号肽，因而不是分泌蛋白。
- (4) 亚细胞定位发现，目的蛋白有可能位于细胞质基质中。
- (5) 结构域预测可知，该结构域为HLH结构域，很有可能是转录因子，在叶片发育过程中发挥调控作用。



➤ 下一步计划：

1. 制作半薄切片，观察白菜不同发育时期叶片背腹面的结构差异；
2. 克隆目的基因；
3. 观察拟南芥突变体叶片的变化情况，筛选背景一致的纯合突变体，用目的基因转化突变体，制作半薄切片，观察突变体与野生型叶片的差异；
4. 用目的基因转化白菜突变体，制作半薄切片，观察突变体与野生型叶片的差异；
5. 对目的基因与调控通路上其他基因的相互作用进行研究，以期了解目的基因的作用机制。



参考文献:

- [1] Catherine A. K. and Marja C. P. Timmermans.
Signaling Sides: Adaxial–Abaxial Patterning in Leaves [J]. *Current Topics in Developmental Biology*. 2010.91: 141-168.
- [2] Nathaniel P. Hawker and John L. Bowman. Roles for Class III HD-Zip and KANADI Genes in Arabidopsis Root Development[J]. *Plant Physiology*, 2004, 135: 2261-2270
- [3] Anat Izhakia and John L. Bowman. KANADI and Class III HD-Zip Gene Families Regulate Embryo Patterning and Modulate Auxin Flow during Embryogenesis in Arabidopsis[J]. *The Plant Cell*, 2007, 19: 495–508
- [4] "KANADI regulates organ polarity in Arabidopsis."
Kerstetter R.A., Bollman K., Taylor R.A., Bomblies K., Poethig R.S.
Nature 411:706-709(2001)
- [5] "Establishment of polarity in lateral organs of plants."
Eshed Y., Baum S.F., Perea J.V., Bowman J.L.
Curr. Biol. 11:1251-1260(2001)



致谢



感谢罗老师把我们带进生物信息的殿堂，在一个学期的学习过程中，我们收获了知识，掌握了方法，学会了思考，老师认真负责的态度将不断激励着我们在科研领域努力学习，勤于钻研，积极探索生命的奥秘。

感谢各位组员的配合，大家在一起学习与讨论，对这学期所学知识有了更加深刻的理解，相信会对大家将来的科研工作有很大的帮助。



Thank you ...

