



甘蓝型油菜PDF1基因的生物信息学分析

Bioinformatic Analysis of Brassica napus PDF1 gene

报告人：桑世飞

组长：周梦妍

组员：周梦妍 桑世飞 马小良 朱晓凤

内容



- 一、PDF1基因的基本信息
关于PDF1基因的相关介绍。
- 二、PDF1同源基因的寻找
根据拟南芥PDF1基因通过序列比对相似性获得。
- 三、甘蓝型油菜PDF1基因的生物信息学分析
 - a、同源序列比对
 - b、基因结构的分析
 - c、PDF1蛋白多肽序列的获取
 - d、PDF1蛋白疏水性分析
 - e、PDF1TMHMM程序进行跨膜分析
 - f、PDF1TargetP程序做亚细胞定位
 - g、SOPMA 对其编码蛋白的二级结构进行预测

一、PDF1基因的基本信息

- 研究表明，拟南芥PDF1的甘蓝型油菜同源基因 BnPDF1在油菜可育株和不育株中的表达具有显著差异。通过荧光定量PCR的方法，对该油菜基因在根、茎、叶、花、雌蕊、雄蕊、角果中的表达模式进行了分析，显示它是一个可育株雄蕊高量表达基因。BnPDF1在油菜雄蕊中的表达和花药育性显著相关。
- PDF1，一个PCD抑制因子，在植株生长、发育中起到重要调控作用。通过对转基因拟南芥进行了深入的细胞学分析，发现PDF1基因是一个细胞程序性死亡的抑制因子，它通过抑制PCD的异常启动来确保细胞分化、分裂以及正常发育。抑制该基因的表达会诱导PCD的提早发生，从而导致细胞败育，造成植株的雄性不育和矮化。

二、PDF1同源基因的寻找



• a、直系同源与旁系同源

- 直系同源(orthologous) 指的是不同物种之间的同源性，例如蛋白质的同源性，DNA序列的同源性。
- Orthologs是指来自于不同物种的由垂直家系（物种形成）进化而来的蛋白，并且典型的保留与原始蛋白有相同的功能。
- 旁系同源（Paralogs）是那些在一定物种中的来源于基因复制的蛋白，可能会进化出新的与原来有关的功能。

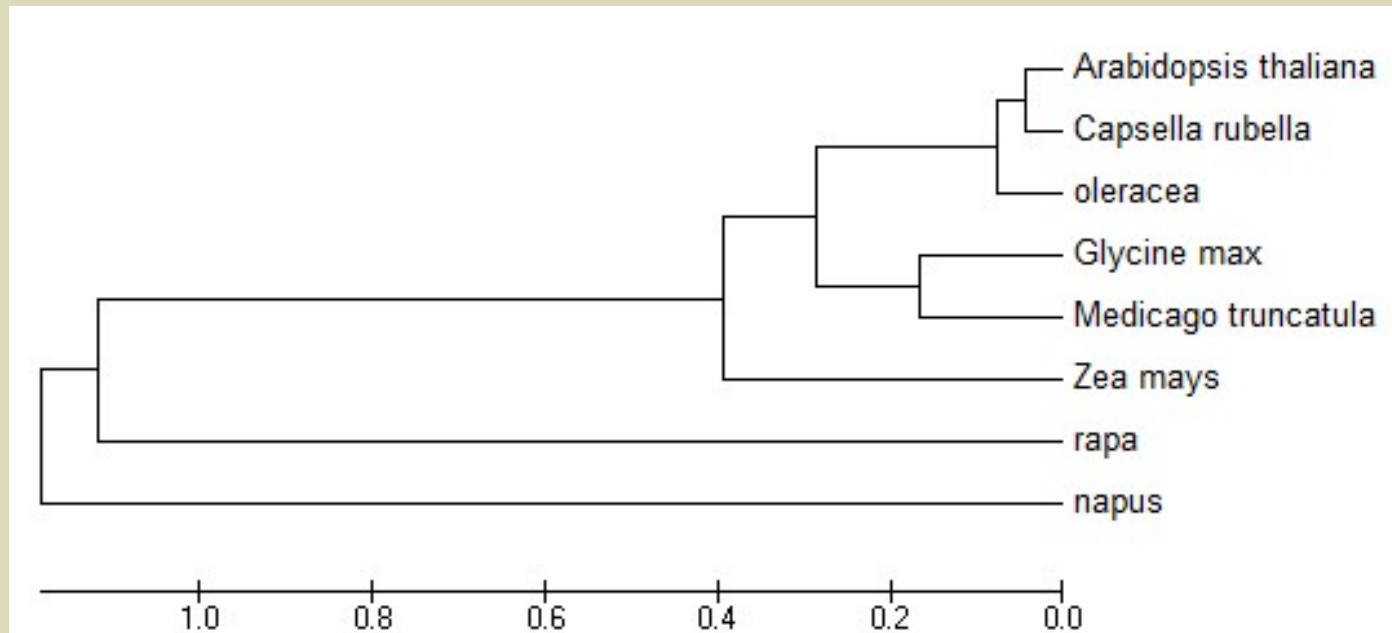
b、 PDF1同源序列的寻找



- 双向 blast (bidirectional best hit method in BLAST)
- 根据拟南芥 Tair 网上提供的拟南芥 PDF1 (AT2G42840.1) 的核苷酸序列，同时根据油料所实验室甘蓝型油菜全基因组测序信息进行比对，获得到与其相似性很高的序列片段。



- 进化树的构建



三、PDF1基因的生物信息学分析



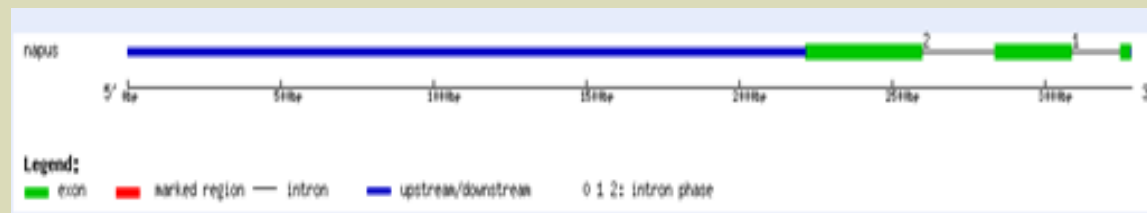
- a、甘蓝型油菜与拟南芥PDF1基因同源序列比对

Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
1222	2748.0	833/1222 (68.2%)	833/1222 (68.2%)	314/1222 (25.7%)

b、基因结构的分析

- 为了对甘蓝型油菜中 PDF1 基因进行分析，采用网上在线基因结构显示系统(Gene Structure Display Server, GSDS, <http://gsds.cbi.pku.edu.cn/>)，对甘蓝型油菜PDF1基因的 CDS 和基因的全序列比较的基础上绘制出它们的基因结构示意图如下：



- 从基因结构图可以看出甘蓝型油菜的基因，基因含有三个外显子（绿色），两个内含子和 5'和 3'的非编码区（蓝色）。

c、甘蓝油菜PDF1蛋白序列的获取

- 在 (<http://www.expasy.org/genomics>) 提供了EMBOSS translation tools这样一个在线的序列翻译程序，利用该程序将获得的甘蓝型油菜PDF1序列即可得到相应的蛋白多肽序列：

```
>napus_1
MTCVYCKVVPTNEVPSPKETDLPLNAYMLHNLAHIELNAIDLAWDTVARFSPFFDVLGRK
FFDDFAHVADDESRRHFMWYGDMPANNLLMRECENTSNNVAARLAVIPLVQEARGLDAGPR
LVKRLMGFGDHRTSKIVAKIAEEGIAHVAVGVDFWFLSVCQKMNRAPCPTFKGTDKYDPSC
ATEADEGGNKQGDKEKLSAVYDRLTHLISMETSRLCYGDSFVYDPSKTLRERFKKWLQTH
SILYGGKEEWMLRFGIYQSNLQLIDYINSLHLPFKLTDNRFADMTNAEFKAHFLGLNTST
LRLNTDQRPVCDAVGNVPASVDWRKEGAVTPVRNQKCGGCWAFATVAAIEGINKIKTGN
LVSLSEQQILIDCDISAFNKGCSGGLLTAYEYLI PNGGIVTEADYPYTAIQGTCDQESQ
NKG FVEIILIME*
```

- 为了检验EMBOSS translation tools的可靠性，将所获得的AA序列与实验室所拿到的序列Needle比对结果如下：

Pairwise Alignment Result

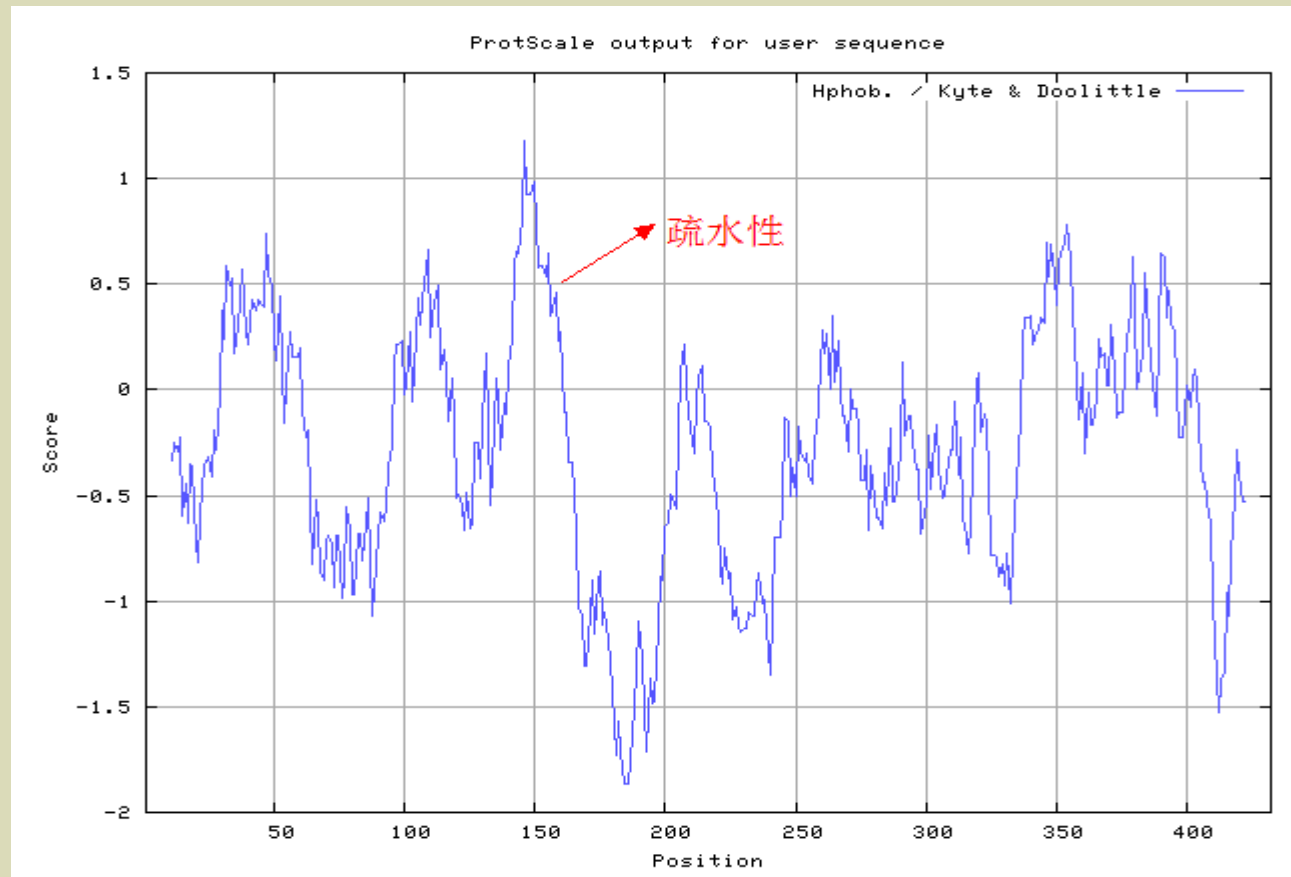
LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
433	2307.0	432/433 (99.8%)	432/433 (99.8%)	1/433 (0.2%)

napus_1	1	MTCVYCKVVPTNEVPSPKETDLPLNAYMLHNLAHIELNAIDLAWDTVARF	50
napus	1	MTCVYCKVVPTNEVPSPKETDLPLNAYMLHNLAHIELNAIDLAWDTVARF	50
napus_1	51	SPFFDVLGRKFFDDFAHVADDES RHF MWYGDMPANNLLMRECENTSNNVA	100
napus	51	SPFFDVLGRKFFDDFAHVADDES RHF MWYGDMPANNLLMRECENTSNNVA	100

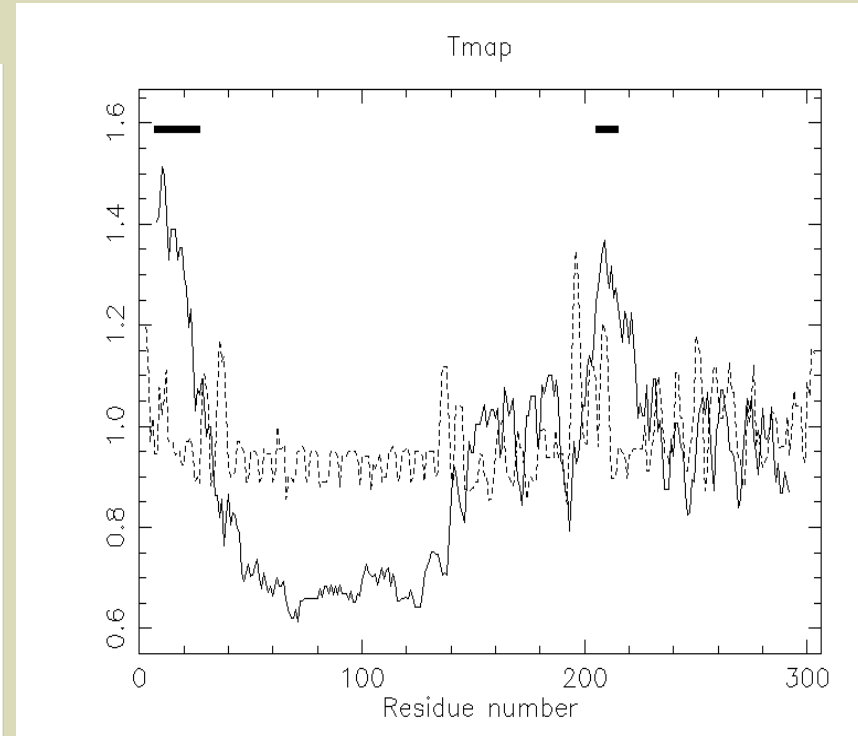
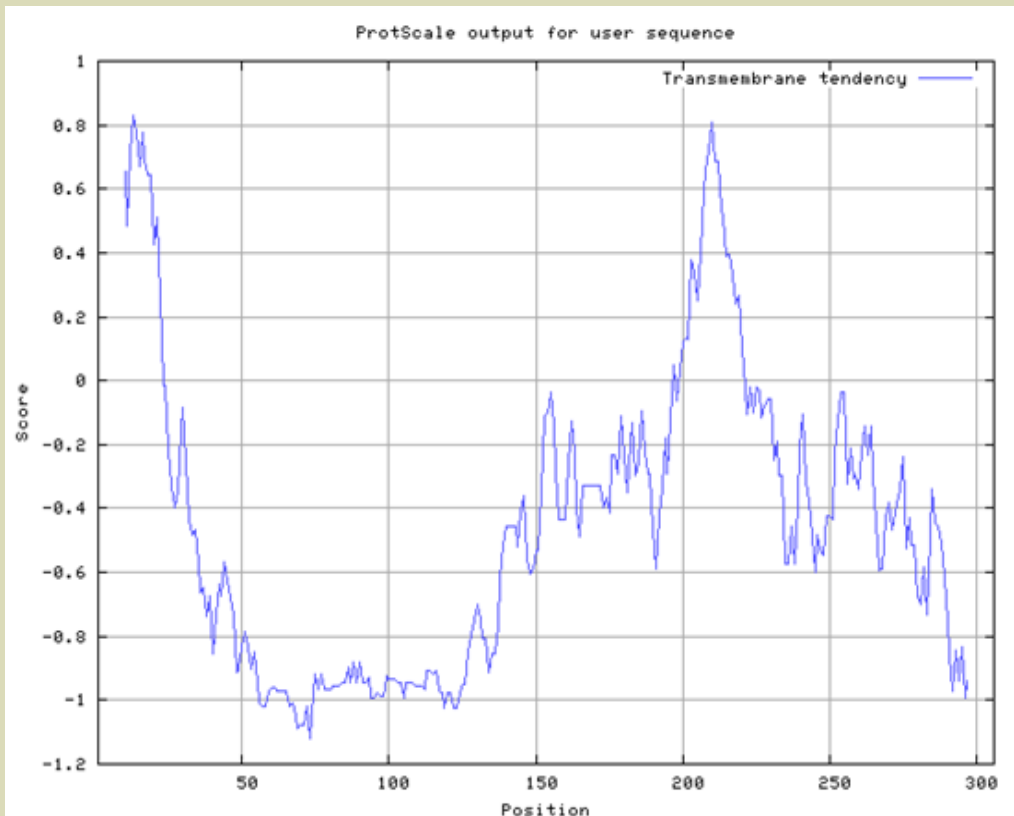
d、PDF1蛋白亲疏水性分析

- 蛋白质的亲疏水性与其翻译的蛋白在亚细胞中的定位有很大关系，若亲水性较高，表明翻译的蛋白在膜上的可能性较高。通过在线软件 ProtScale (<http://www.expasy.ch/tools/protscale.html>) 对甘蓝型油菜 PDF1 基因进行分析。Kyte & Doolittle 计算疏水性，得0分以下表示亲水，0分以上表示疏水。ProScale 的默认值为 9，这里将 Kyte & Doolittle 计算时的 window size 调为 21 左右就可以同时分析是否有跨膜结构区段。结果下页：

- 甘蓝型油菜的 PDF1基因编码的氨基酸，亲水性和疏水性蛋白延0轴线较均匀的分布，两序列大部分的氨基酸在0轴线的下方，因此大部分的氨基酸属于亲水氨基酸。



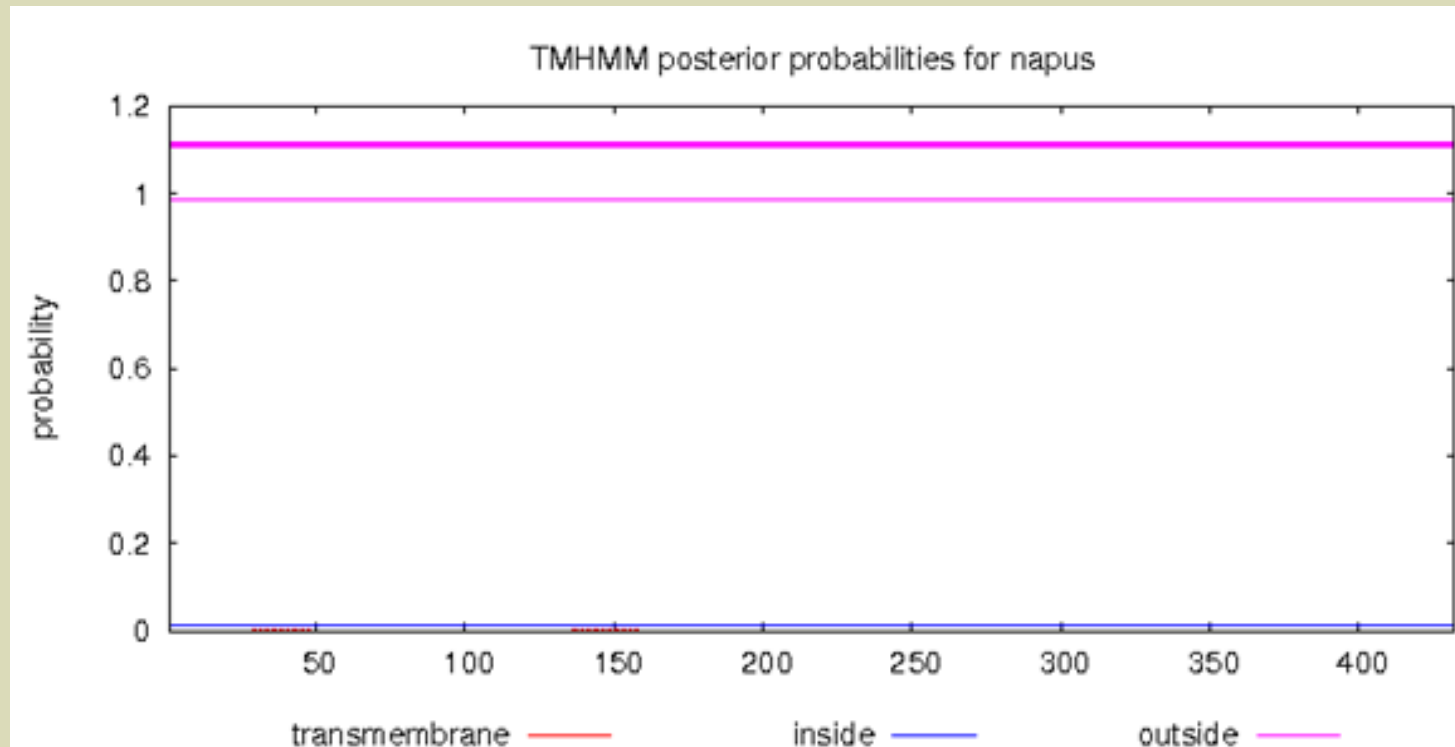
拟南芥PDF1蛋白序列亲水性分析



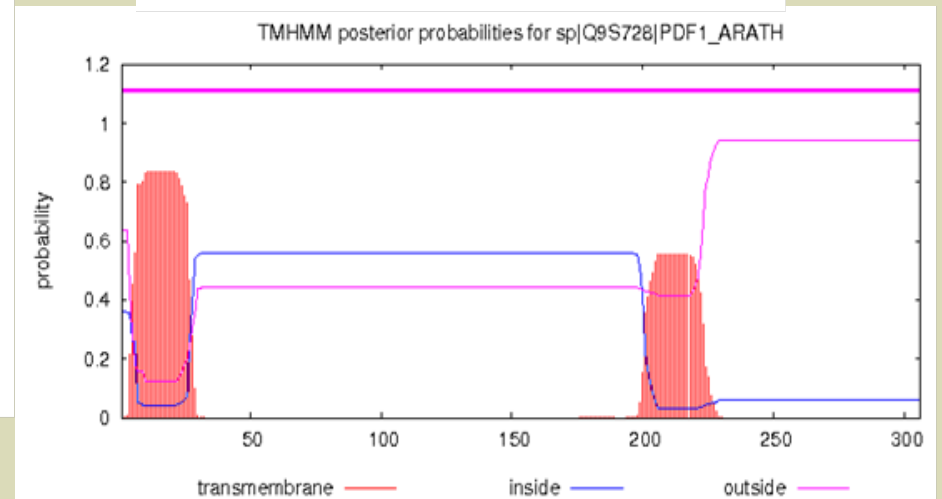
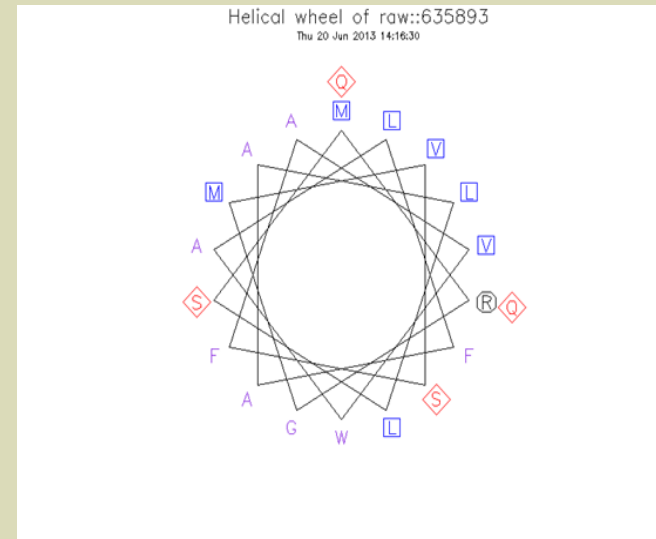
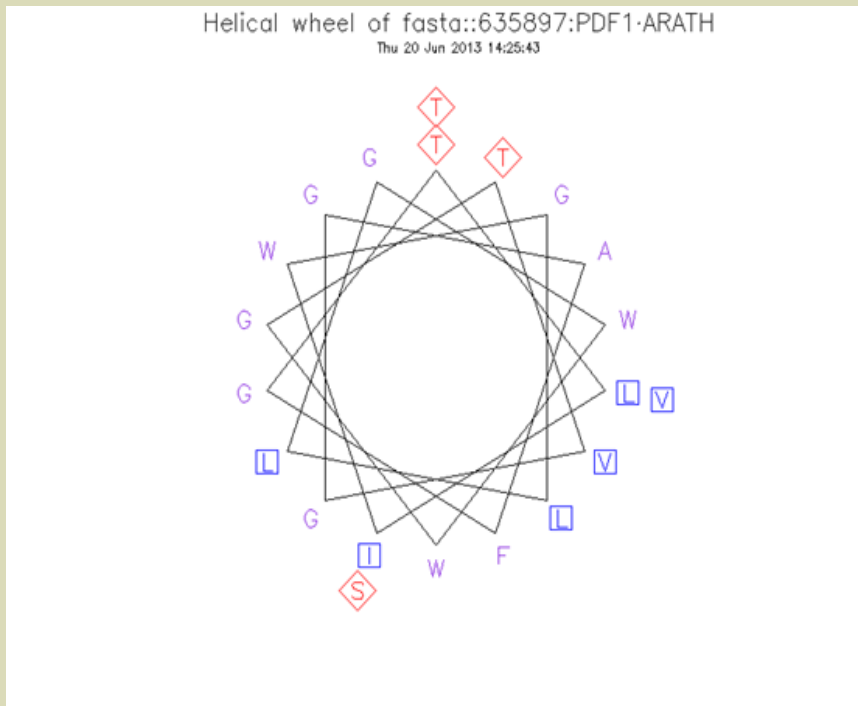
e、PDF1蛋白的跨膜分析

- 在上述分析基础上，我们利用 TMHMM 在线软件（<http://www.cbs.dtu.dk/services/TMHMM/>）对甘蓝型油菜PDF1 蛋白的跨膜结构域进行了分析。TMHs 的数量表示有多少个区段是跨膜的；AAs在TMHs 中的数量表示有多少氨基酸参与了跨膜；60AAs 的期望值表示对N端 60个氨基酸是否为信号肽的预测，如果数值高（大于 10），很可能表示 N 端包含着一段信号肽；在细胞质侧的预测值是N-in,范围为 0~1。

- 结果表明，甘蓝型油菜蛋白序列的 AAs 值为0.04表明序列没有信号肽段。甘蓝型油菜的PDF1蛋白序列的N-in值为0.01较低，甘蓝型油菜的PDF1蛋白可能是粘连在细胞膜上。



用TMHMM分析拟南芥PDF1跨膜



f、PDF1蛋白的亚细胞定位

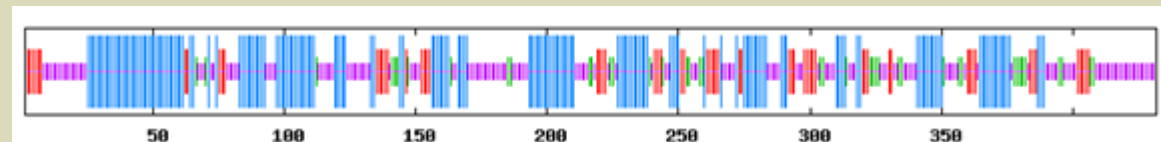
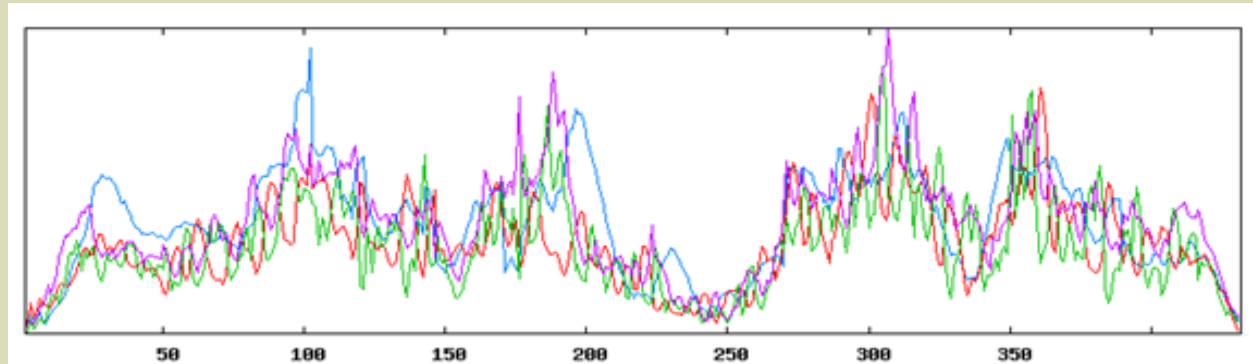
- 为了进一步明确PDF1蛋白的具体位置，做更进一步的定位，利用在线程序TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) 做亚细胞定位。

```
Number of query sequences: 1
Cleavage site predictions not included.
Using PLANT networks.
```

Name	Len	cTP	mTP	SP	other	Loc	RC
napus	432	0.058	0.112	0.127	0.902	_	2
cutoff		0.000	0.000	0.000	0.000		

g、SOPMA 对编码蛋白的二级结构进行预测

- 在线程序SOPMA提供了对二级结构预测的功能。



- 应用 SOPMA 对其编码蛋白的二级结构进行预测，我们从预测结果可以看出，甘蓝型油菜的 α 螺旋、延伸链、 β 转角和不规则卷曲所占全长的比例。

```
Sequence length :    432

SOPMA :
  Alpha helix      (Hh) :   166 is  38.43%
  310 helix      (Gg) :     0 is   0.00%
  Pi helix        (Ii) :     0 is   0.00%
  Beta bridge     (Bb) :     0 is   0.00%
  Extended strand (Ee) :    64 is  14.81%
  Beta turn       (Tt) :    41 is   9.49%
  Bend region     (Ss) :     0 is   0.00%
  Random coil     (Cc) :   161 is  37.27%
  Ambiguous states (?) :     0 is   0.00%
  Other states    :     0 is   0.00%
```



敬请老师、同学们批评指正！