

# 普通小麦果聚糖-6果糖基转移酶6-SFT的生物信息学分析

小组:G11

汇报人:董艳

小组成员:白娜 王翠红  
董艳 卢家玲



# 提纲

1

研究背景及意义

2

基因序列分析

3

蛋白质序列分析

4

蛋白质功能预测

# 研究背景及意义

- 小麦在中国是主要的谷物和粮食作物,约有2400万公顷,平均产量4.7吨/公顷,总产量约11亿吨.
- **果聚糖**是温带和冷寒带地区碳水化合物的主要储存形式,主要存在禾本科作物小麦、大麦、梯牧草中.果聚糖有多种生物学功能.

## 果聚糖代谢的生理作用

- 作为储藏性碳水化合物
- 调节光合组织中蔗糖库的大小
- 适应低温光合作用
- 对缺氧的适应性
- 适应低温、增强抗旱性
- 有效减少和降低结肠癌,心血管疾病和骨质疏松症等发病几率。



食用含果聚糖的食物,可以选择性地促进双歧杆菌的生长,减少与肿瘤发生有关的物质产生。

# 6-SFT的催化机制

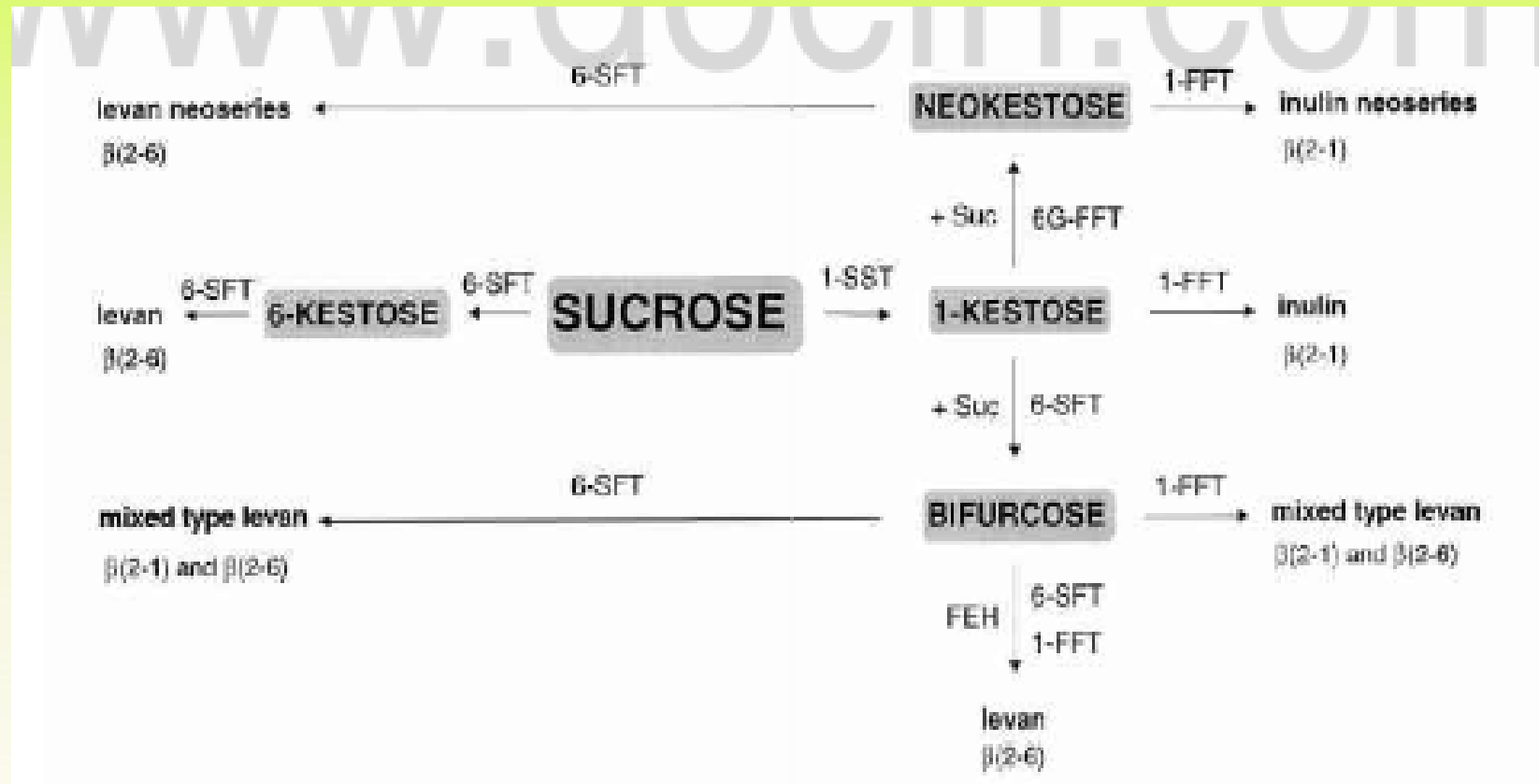

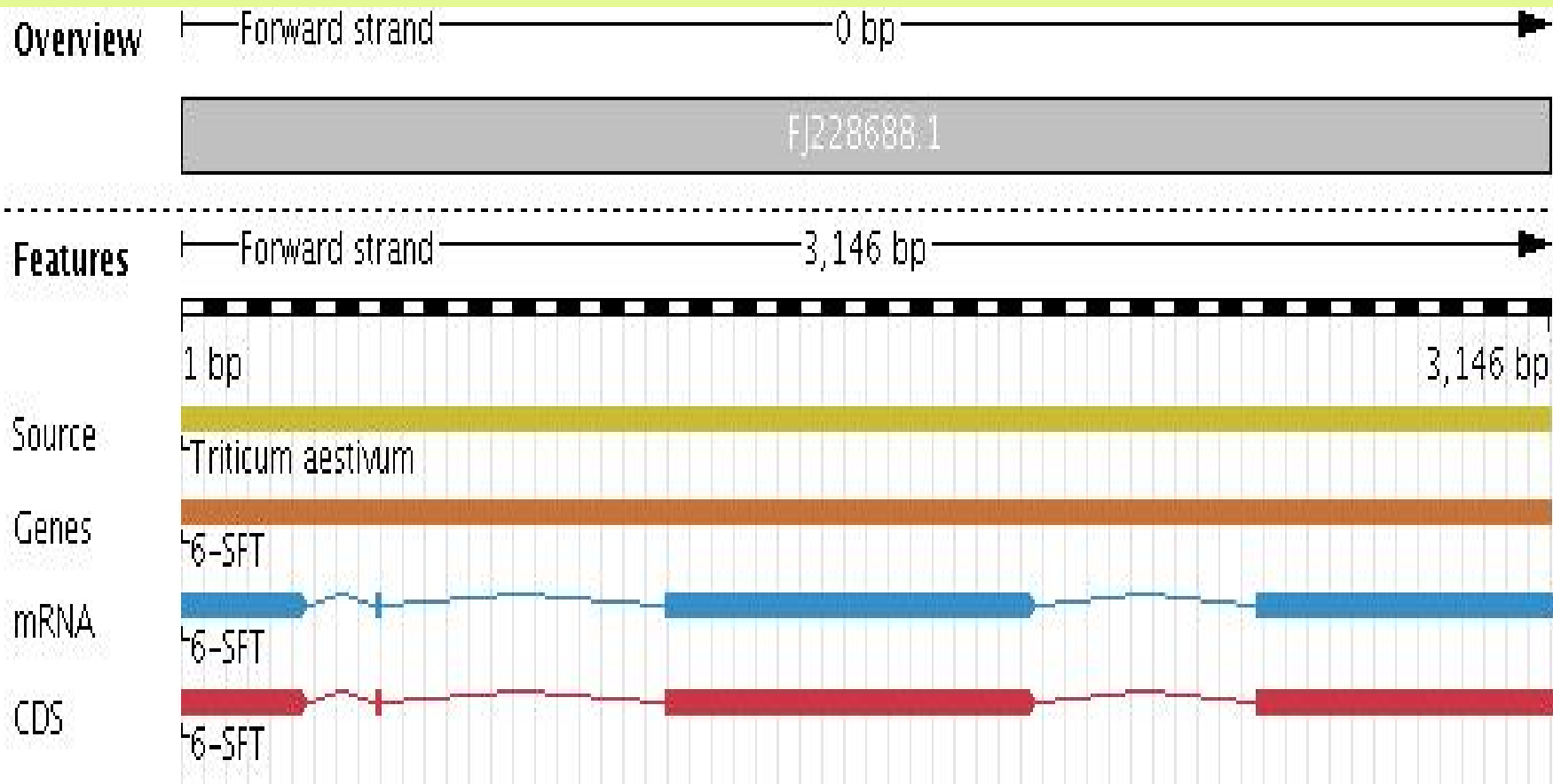


图 1 Vijn和 Smeekens (1999) 的果聚糖代谢模型

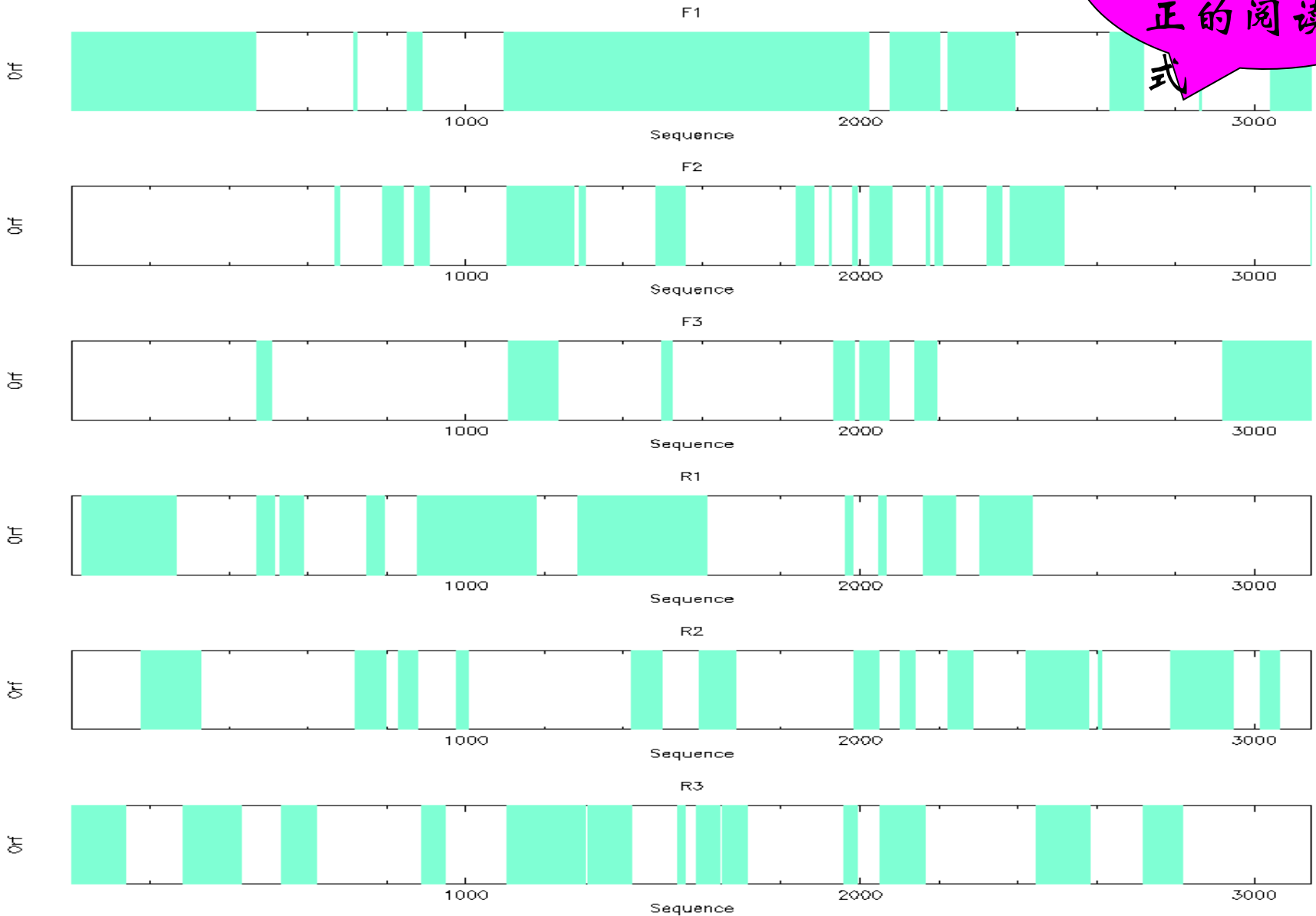
- 
- 蔗糖-果聚糖-6-果糖基转移酶 (Sucrose: fructan 6-fructosyltransferase, 6-SFT), 是果聚糖合成过程中的关键调节酶. 在进化上起源于液泡转化酶, 属糖基水解酶32家族 (glycosyl hydrolase family 32, GH32), N末端为 $\beta$ -螺旋模型, 由五个刀片状结构组成, C末端有一个 $\beta$ -夹心模型.
  - 本实验拟研究6-SFT的结构、功能, 从而为克隆6-SFT基因奠定基础.

# 6-SFT 已公布的信息



# plotorf 对开放阅读框的预测

F1 阅读框  
最有可能是  
正的阅读方  
式





# -coderet 提取CDS序列

□ >fj228688\_cds\_1

```
atggggtcacacggcaagccaccgctaccgtacgcgtacaagccactgcctccggcgccgctggacgccgacggtgagcggaccggctggacg
aggtggcgcgtgtgcgccaccgtgctgacggcctcggccatggtggtggtggtggtcggcgccacgctcctggcaggggttcaggggtggaccaggccgt
cgacgaggaggccgcggccgggtcccgtggagcaacgagatgctgcagtggcagcgcagtggctaccattccagacggccaagaactacatgacg
gatcccaacggtcttatgtactacaatggatggtaccacatgctctccagtacaaccgggtgggcaccgattgggacgacggcatggagtggggccatgc
cgtgtctcggaaacctgtcacgtggcgcacccttctattgccatggtggtgaccagtggtagacatcctgggggtccttcgggctctatgacgggtgta
cccaatggcacggttatcatgatctacacggggggccaccaacgcctctgccattgaggtgcagtgcacgccacccccgccgacccaacgacccctcc
tccgccgctggaccaagcaccgcccaacccgctcatttgctcggccggggatcggcaccaggatttcgagaccgatgactgcttggtacgatga
atctgatgacacatggcgcaccctccttgggtccaaggatgaccacgacggtcaccacgatgggatcggcatgatgtacaagaccaaggacttcctaact
acgagctcatcccgggtatcttgcacgagtcaccgcaccggcgagtgaggagtgcattgacttctaccctgtcggccacagaagcaacgacaactcatcg
gagatggtgacggttgaaggcgagcatggacgacgaacggcacgactactactcgttaggcacgtacgactcggcagcaaacacgtggacgccgatc
gacccggagctcgacttggggatcgggctgagatacactgggtaagttttatgctccacctcgttctatgatccggcaagaagcggcgcgtgtgat
ggggtacgtcggcgaggtcgactccaagcgggctgatgtcgtgaagggtgggctcgattcggcagttccaaggacgattgctctggacgagaagac
ccggacgaacctcctccttggcccgtggaggagattgagacctccgctcaacgccaccgaacttagcgacgtcaccttaacaccggctcggctcacc
catacccgcctccccaaggcactcagctcgacatcaggcaacttccaccttgatgcttctgccgtcgtgcctcaatgaggccgatgtggggtacaact
gcagcagcagcggcggtgctgtaaccgcggcgcgctaggcccccttggcctcctcgtctcgtgctggtgaccgccgtggcgagcaaacggcggtgt
atttctacgtgtctagggggctcgacggaggcctccataccagcttctgccaagacgagttgcggtcgtcacgggccaaggatgtgacgaagcgggtgatt
gggagcacggtgccggtgctcgacggcgaggcttctcgtatgagggtgcttggaccactccatcgtgcagggcttcgcgatgggcgggaggaccac
gatgacgtcgcgggtgtaaccgatggaggcctatcaggaggcaaaagtgtacttgtcaacaatgcgaccgggtgccagcgtcatggcggaaggctcgtc
gtgcacgagatggactcagcacacaaccagctctccaatatggacgatcactcgtatgttcaatga
```

□ 1878个碱基

# -coderet提取蛋白序列

>fj228688\_pro-1

```
MGSHGKPLPYAYKPLPSGAAVDADGERTGWTRWRVCATVLTASAMVVVVVGATLLAGFR
VDQAVDEEAAAGFPWSNEMLQWQRSGYHFQTAKNYMSDPNGLMYYNGWYHMFQYNPVG
DWDDGMEWGHAVSRNLVTWRTLPIAMVADQWYDILGVLSGSMTVLPNGTVIMIYTGATNA
SAIEVQCIATPADPNDPFLRRWTKHPANPVIWSPPGIGTKDFRDPMTAWYDESDDTWRTL
LGSKDDHDGHHDIAMMYKTKDFLNYELIPGILHRVQRTGEWECIDFYVPVGHRSNDNSSE
MLHVLKASMDDERHDYYSLGTYDSAANTWTPIDPELDLIGLRYDWGKFYASTSFYDPAK
KRRVLMGYVGEVDSKRADVVKGWASIRSVPRITIALDEKTRTNLLLWPVEEITLRLNATE
LSDVTLNTGSVIHIPLRQGTQLDIEATFHLDASAVAALNEADVGYNCSSSGGAVNRGALG
PFGLLVLAAGDRRGEQTAVYFYVSRGLDGGLHTSFCQDELRSSRAKDVTKRVIGSTVPVL
DGEAFSMRVLVDHSIVQGFAMGGRTTMTSRVYPMEAYQEAKVYLFN NATGASVMAERLVV
HEMDSAHNQLSNMDDHSYVQ
```

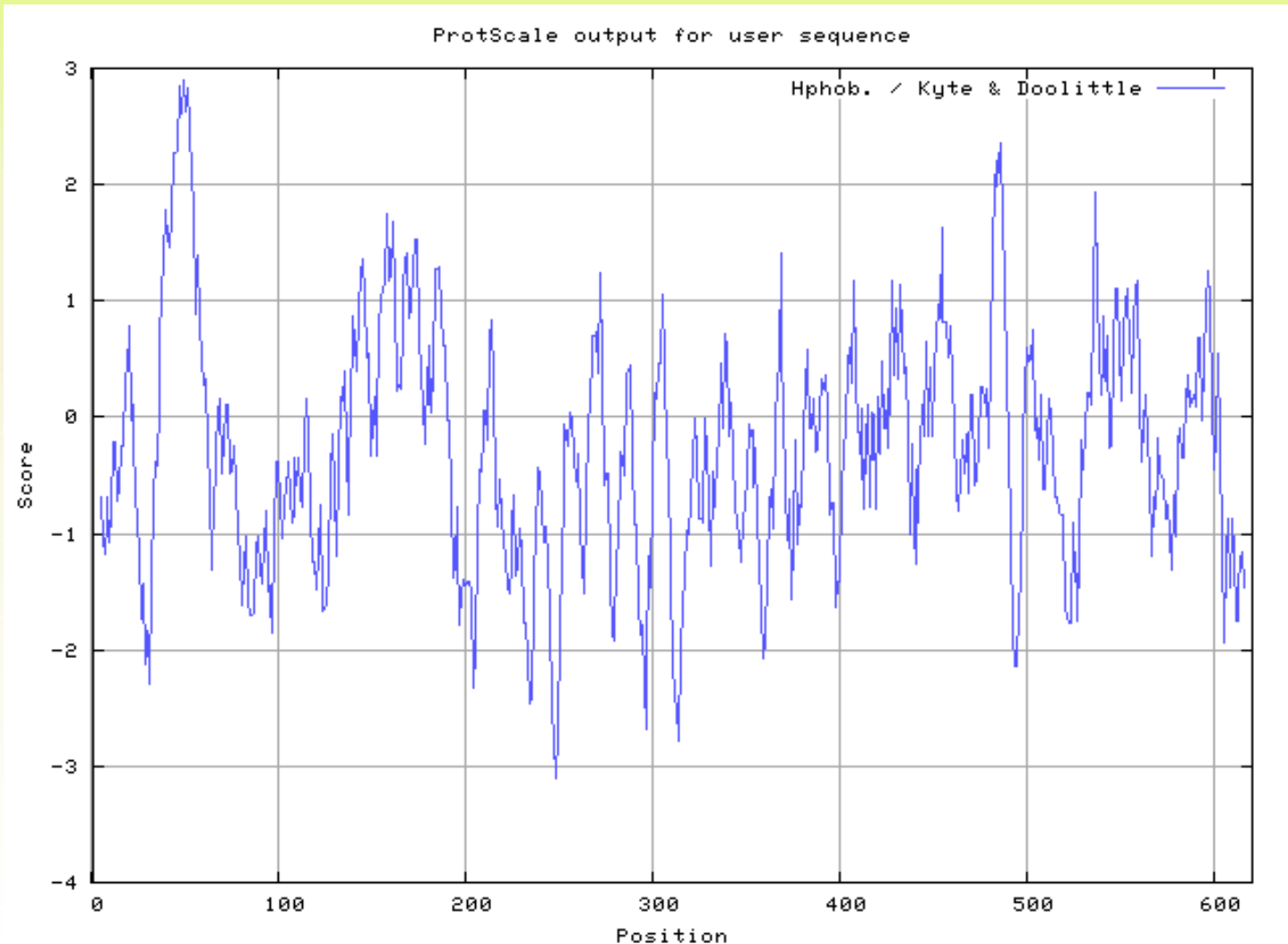
□ 620个AA,与Uniprot中公布的一致

# 蛋白质序列分析

- 1、亲疏水性分析
- 2、跨膜区预测
- 3、信号肽预测
- 4、亚细胞定位
- 5、二级结构分析

# 1、亲疏水性分析

- 通过 (<http://web.expasy.org/protscale/>) 中的Kyte对该蛋白质进行亲疏水性分析, 其结果如下:



# 等电点分析

Ala	(A)	54	8.7%
Arg	(R)	35	5.6%
Asn	(N)	24	3.9%
Asp	(D)	48	7.7%
Cys	(C)	5	0.8%
Gln	(Q)	16	2.6%
Glu	(E)	28	4.5%
Gly	(G)	52	8.4%
His	(H)	19	3.1%
Ile	(I)	22	3.5%
Leu	(L)	49	7.9%
Lys	(K)	18	2.9%
Met	(M)	23	3.7%
Phe	(F)	18	2.9%
Pro	(P)	29	4.7%
Ser	(S)	42	6.8%
Thr	(T)	42	6.8%
Trp	(W)	18	2.9%
Tyr	(Y)	28	4.5%
Val	(V)	50	8.1%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%

等电点PI=5.23

带正电的氨基酸:

(Asp+Glu)=76

带负电的氨基酸:

(Arg+Lys)=53

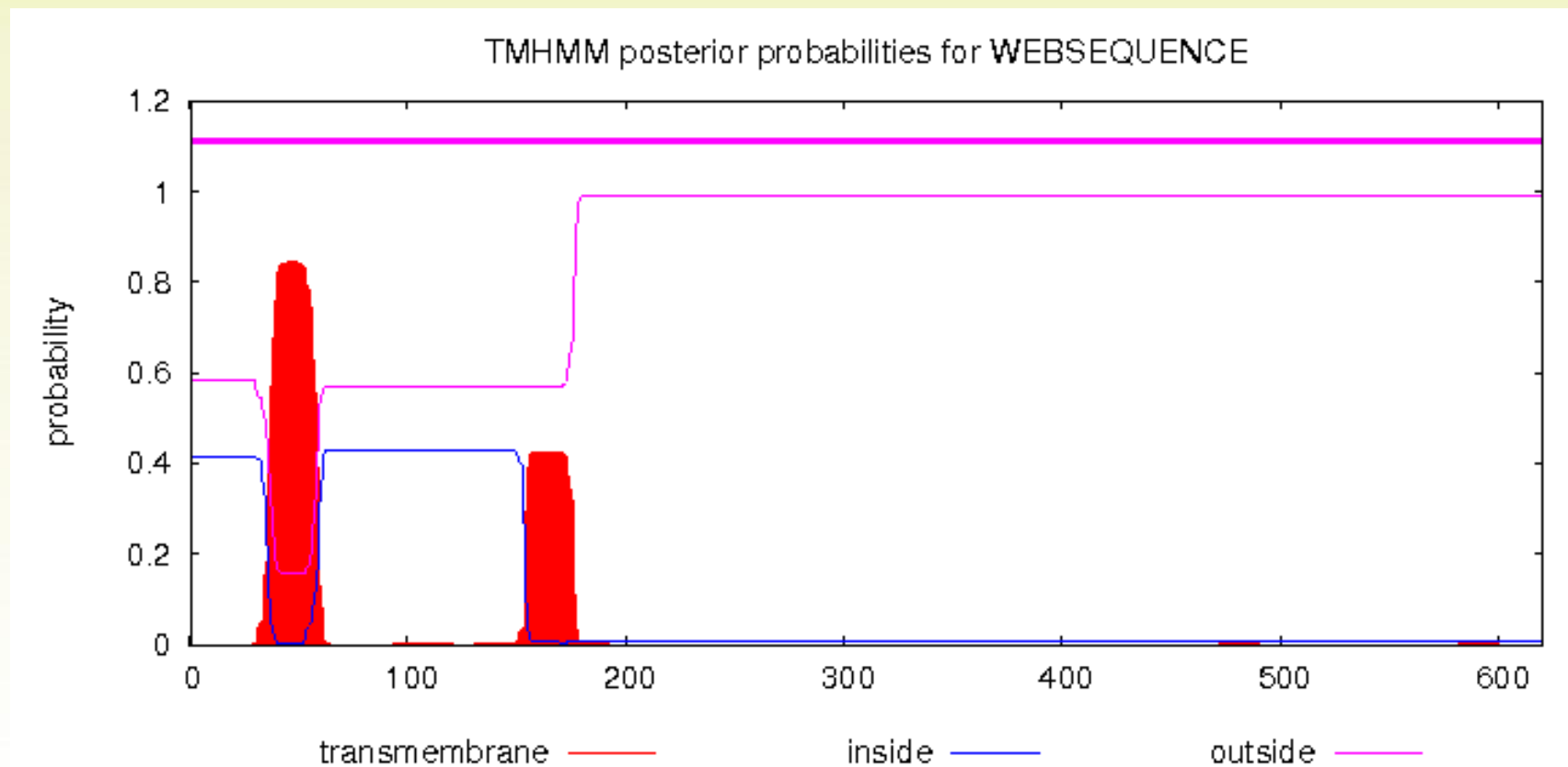
平均疏水指数:

(GRAVY) = -0.288

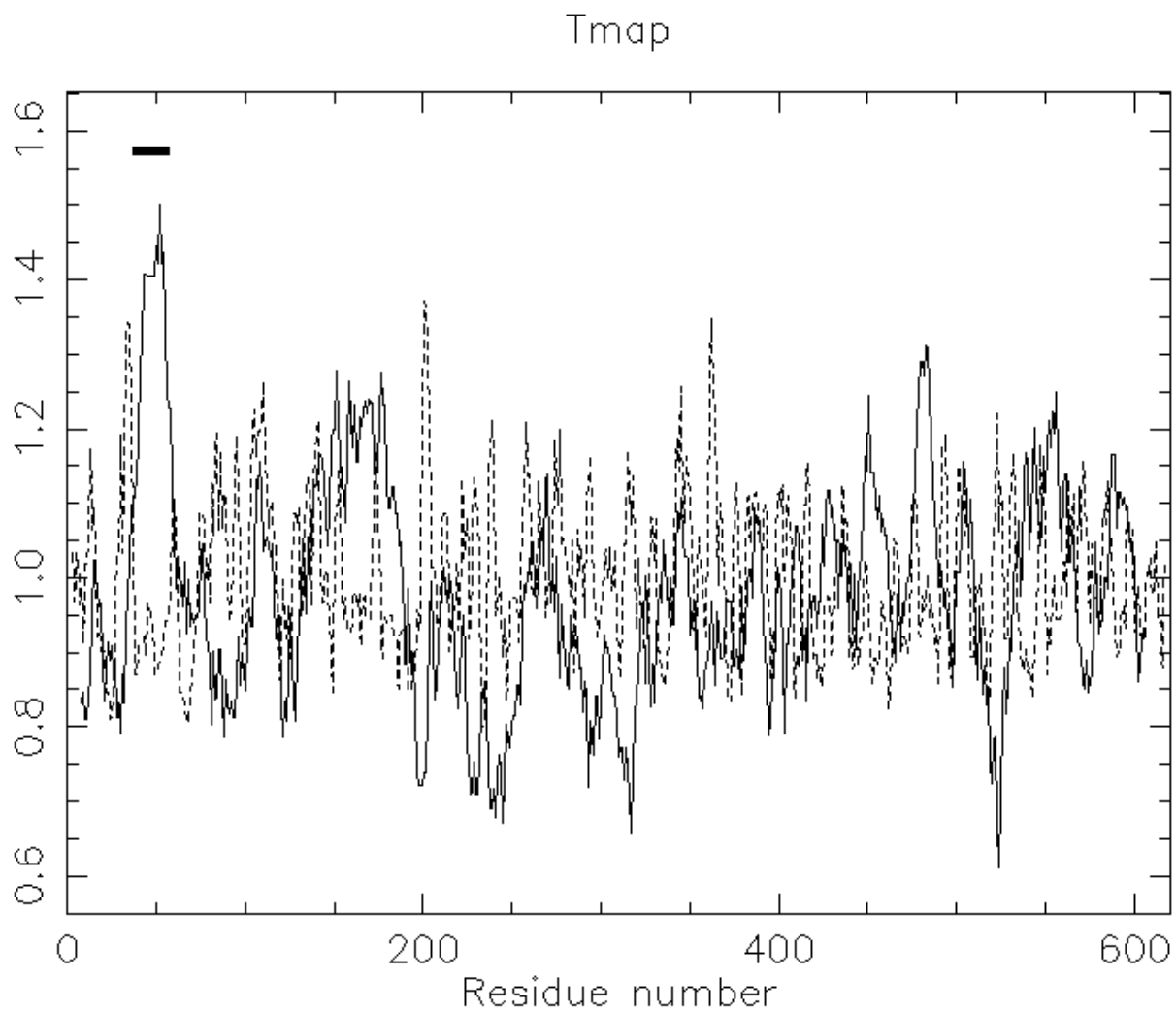
## 2、跨膜区预测

采用两种不同的方法分别预测跨膜螺旋

1 TMHMM Server v. 2.0预测到一个跨膜螺旋

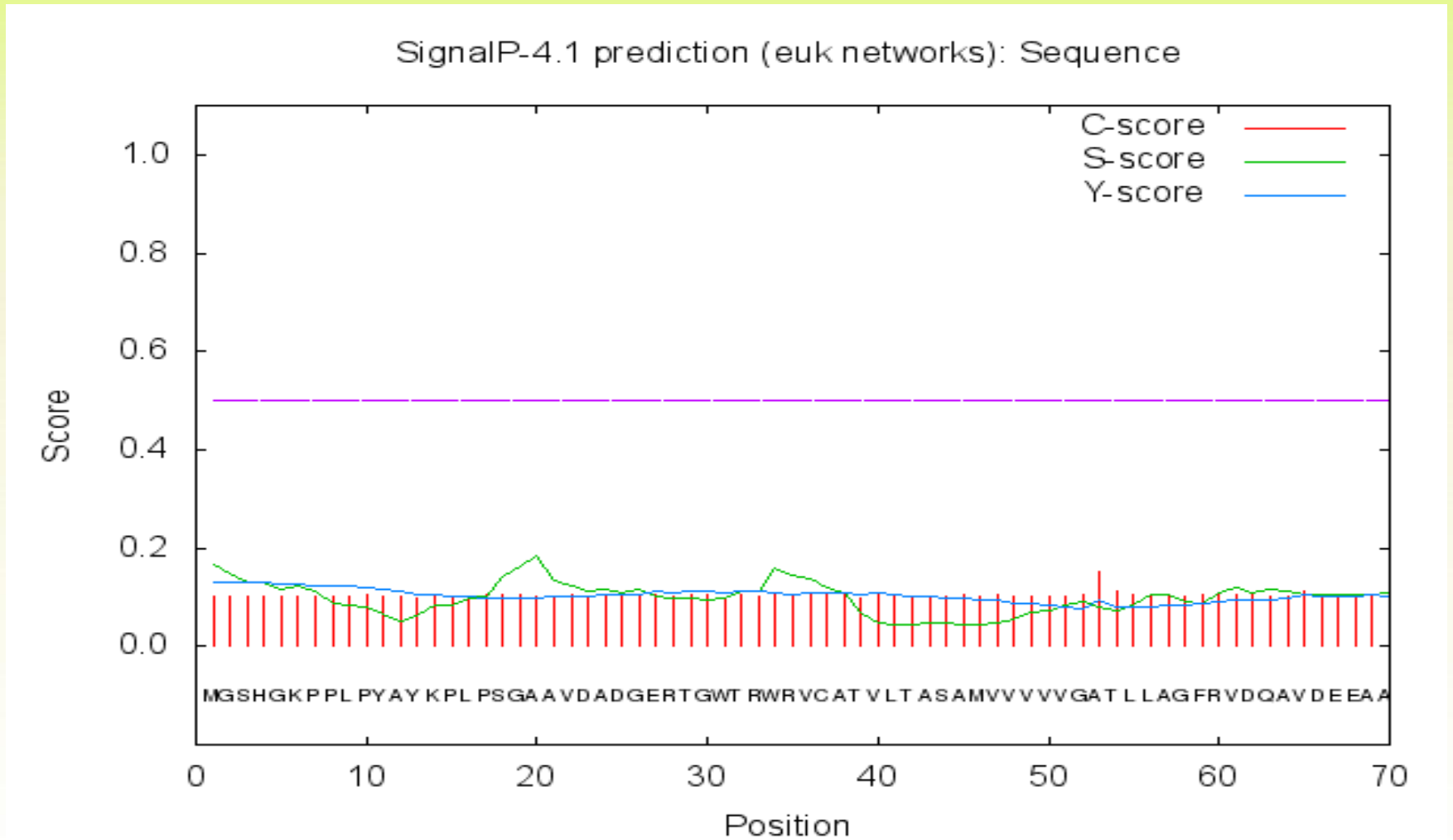


## 2 用weblab中工具tmap进行分析



# 3、信号肽预测

- 使用SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) 对该蛋白序列进行信号肽预测, 结果如下图: **没有预测到信号肽**





# 4、亚细胞定位

- 使用TargetP 1.1 Server对该蛋白进行亚细胞定位如下  
(<http://www.cbs.dtu.dk/services/TargetP/>)

## 定位在线粒体上

说明：cTP代表叶绿体  
mTP代表线粒体  
SP代表分泌途径

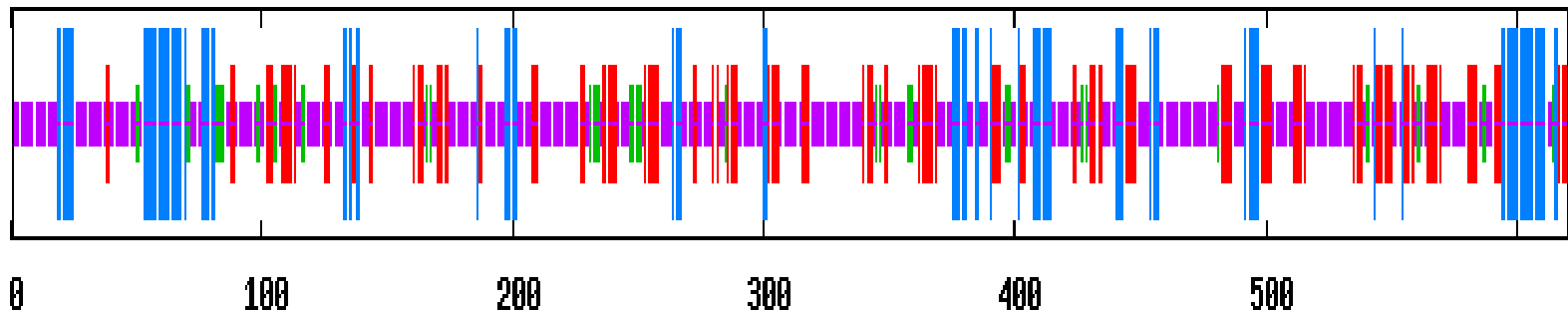
Name	Len	cTP	mTP	SP	other	Loc	RC
Sequence	620	0.033	0.185	0.089	0.873	_	2
cutoff	0.000	0.000	0.000	0.000			

与GH32家族的蛋白绝大多数定位在液泡上不一致

# 4、二级结构预测

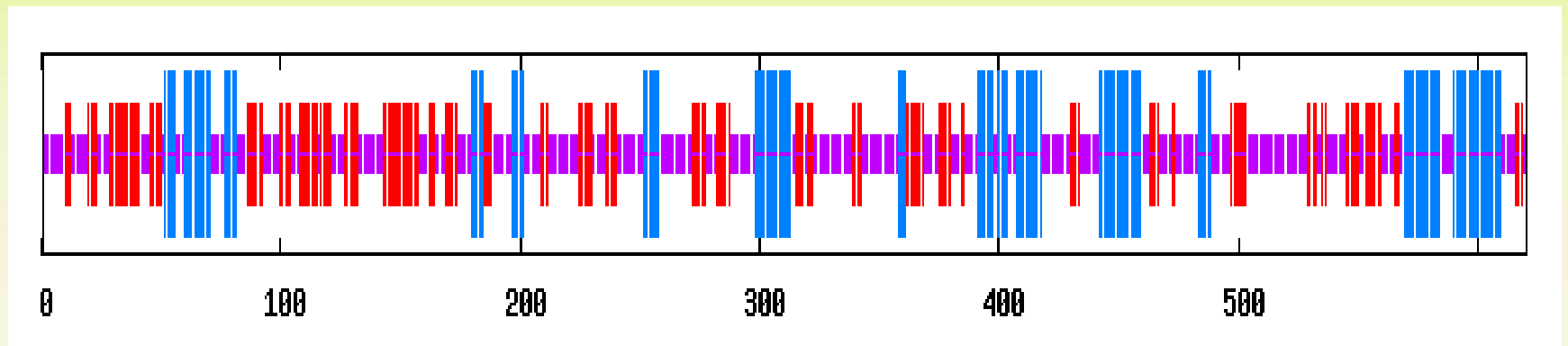
□ 用ExPASy中的SOPMA对该蛋白二级结构如下：

([http://npsa-pbil.ibcp.fr/cgi-bin/secpred\\_sopma.pl](http://npsa-pbil.ibcp.fr/cgi-bin/secpred_sopma.pl))



二级结构类别	AA数目	AA所占比例
Alpha helix (Hh) 蓝色	101	16.29%
Extended strand (Ee) 红色	144	23.23%
Beta turn 绿色	44	7.10%
Random coil (Cc) 紫色	331	53.39%

## □ GOR4 软件预测二级结构如下:



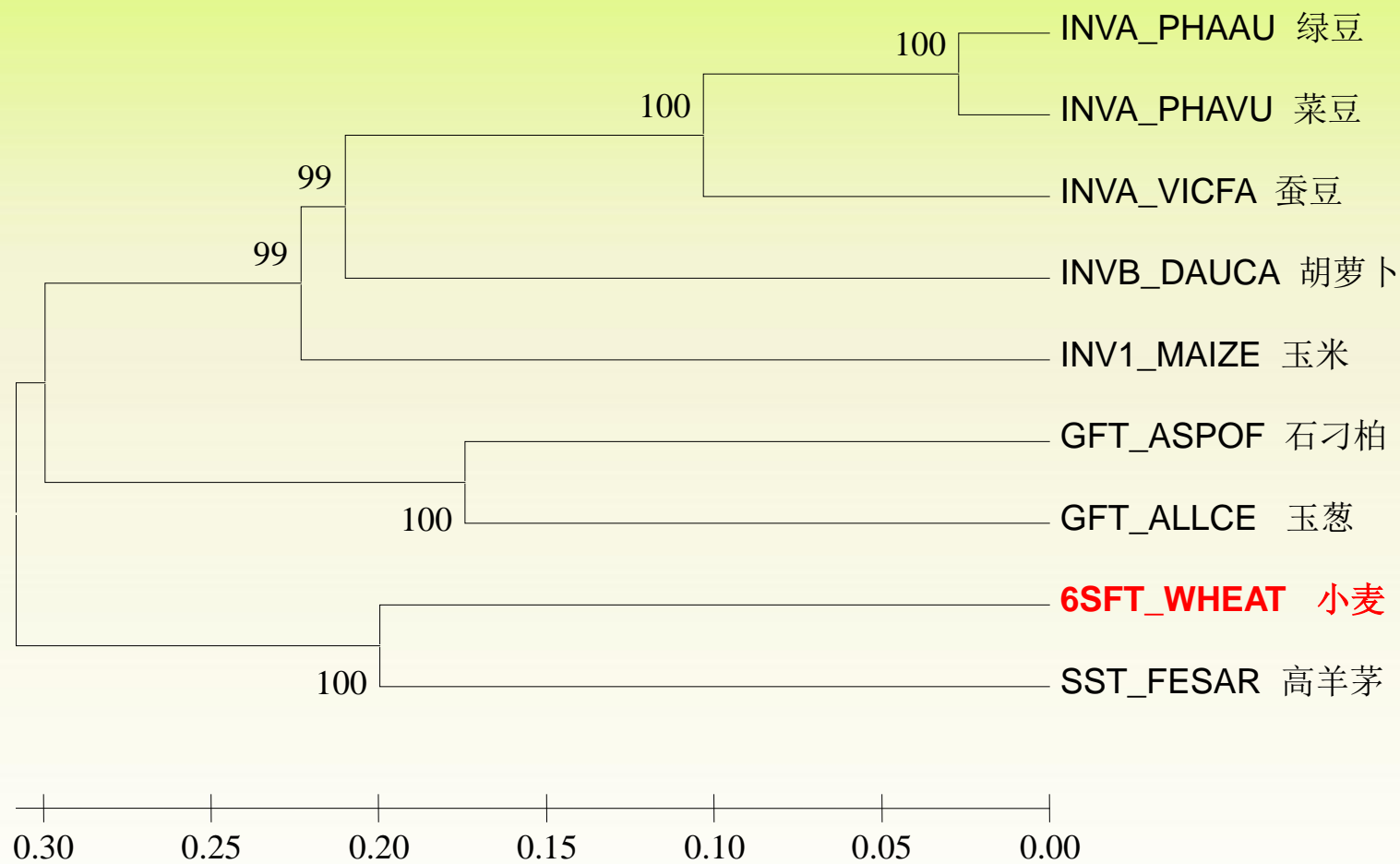
Sequence length : 620AA

- Alpha helix (Hh) : 140 is 22.58%
- Extended strand (Ee) : 176 is 28.39%
- Random coil (Cc) : 304 is 49.03%

使用Blast在Uniprot库中搜索其同源基因,使用ClustalW做多  
序列比对(登录号B6ECPO)

Species/Abbrv	*	*	*	*				*					*	*	*				*	*	*														*	*	*																																		
1. 6SFT WHEAT	H	M	F	F	Q	Y	N	P	V	G	I	D	D	D	G	M	E	F	G	H	A	V	S	R	N	L	V	T	R	T	L	P	I	A	M	V	A	D	Q	W	Y	D	I	L	G	V	L	S	G	S	M	I	V	L	P	N	G	-	I	V	I	M	I	Y	T	G	A	T			
2. SST FESAR	H	L	F	Y	Q	Y	N	P	K	G	D	S	W	G	N	-	I	A	N	A	H	A	V	S	K	D	M	V	N	R	H	L	P	L	A	M	V	P	D	Q	W	Y	D	S	N	G	V	L	T	G	S	I	I	V	L	P	D	G	-	Q	V	I	L	L	Y	T	G	N	I		
3. GFT ASPOF	H	L	F	Y	Q	H	N	P	N	Y	A	Y	W	G	D	-	I	S	W	G	H	A	V	S	R	D	L	L	N	W	F	H	L	P	V	A	V	K	P	D	R	W	Y	D	I	Y	G	V	W	T	G	S	I	I	V	M	P	D	D	G	R	V	V	M	L	Y	T	G	G	T	
4. GFT ALLCE	H	L	F	Y	Q	H	N	K	D	F	A	Y	W	G	N	-	I	T	W	G	H	A	V	S	R	D	L	I	N	W	Q	H	L	P	V	A	V	G	P	D	H	W	Y	D	I	S	G	V	W	T	G	S	I	I	V	V	S	E	D	-	R	V	V	M	L	F	T	G	G	T	
5. INV1 MAIZE	H	L	F	Y	Q	W	N	P	D	S	A	V	W	G	N	-	I	T	W	G	H	A	V	S	R	D	L	L	H	W	L	H	L	P	L	A	M	V	P	D	H	W	Y	D	A	N	G	V	W	S	G	S	A	T	R	L	P	D	G	-	R	I	V	M	L	Y	T	G	S	T	
6. INV A PHAAU	H	F	F	Y	Q	Y	N	P	N	G	A	V	W	G	D	-	I	V	W	G	H	A	V	S	R	D	M	I	H	W	L	H	L	P	L	A	M	V	A	D	Q	W	Y	D	K	Q	G	V	W	T	G	S	A	T	I	L	P	N	G	-	E	I	I	M	L	Y	T	G	S	T	
7. INV A PHAVU	H	F	F	Y	Q	Y	N	P	N	G	A	V	W	G	D	-	I	V	W	G	H	A	V	S	R	D	M	I	H	W	L	H	L	P	L	A	M	V	A	D	Q	W	Y	D	K	Q	G	V	W	T	G	S	A	T	I	L	P	N	G	-	E	I	I	M	L	Y	T	G	S	T	
8. INV B DAUCA	H	L	F	Y	Q	Y	N	P	D	S	A	I	W	G	N	-	I	T	W	G	H	A	I	S	R	D	L	I	N	W	L	H	L	P	F	A	M	G	P	D	Q	W	Y	D	I	N	G	V	W	T	G	S	A	T	I	V	L	P	D	G	-	K	I	V	M	L	Y	T	G	D	T
9. INV A VICFA	H	F	F	Y	Q	Y	N	P	N	G	A	V	W	G	D	-	I	V	W	G	H	A	V	S	R	D	L	I	H	W	L	H	L	P	L	A	M	V	A	D	Q	W	Y	D	S	N	G	V	W	T	G	S	A	T	I	L	P	D	G	-	Q	V	I	M	L	Y	T	G	S	T	

# 构建系统发育树



□ SST-FESAR (高羊茅) Q9FSV7

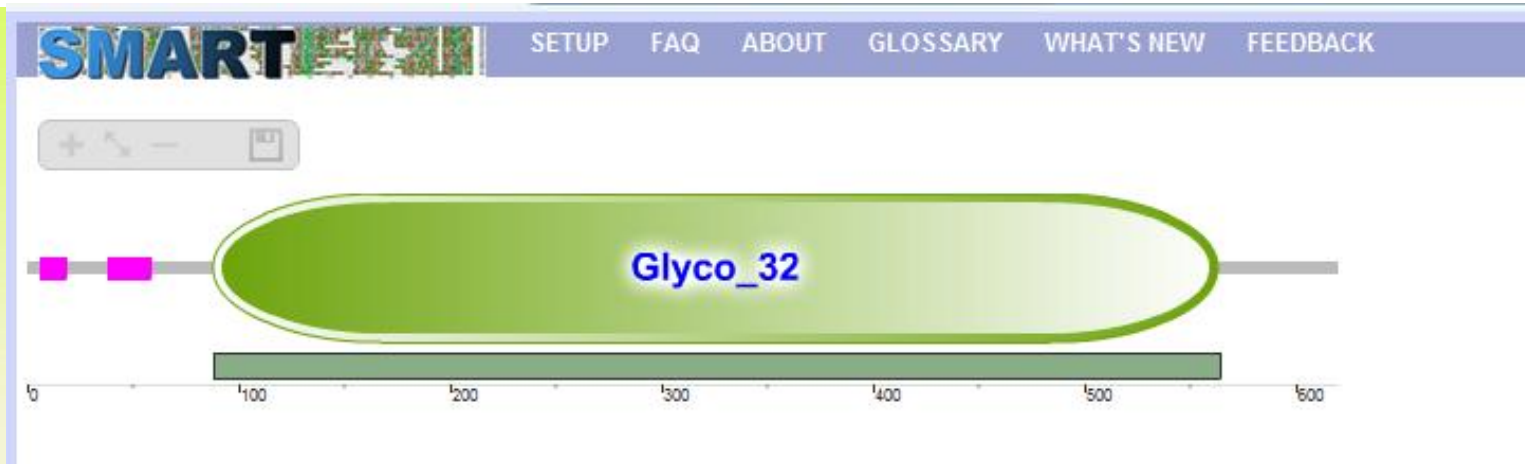
□ 功能已知:

□ 参与果聚糖合成,催化蔗糖转化成果糖和霉菌赤藓糖  
→

□ 有水解酶活性

□ 推测:6-SFT可能也具有这些功能

# SMART 预测蛋白结构域



## Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value	
<a href="#">low complexity</a>	6	17	N/A	▲
<a href="#">low complexity</a>	38	57	N/A	
<a href="#">Glyco_32</a>	88	563	5.65e-222	▼

## Outlier homologues and homologues of known structure: ?

Name	Sequence	Start ▲	End	E-value	
<a href="#">PDB:2AEYJA</a>	2aey	78	596	2e-99	▲
<a href="#">Blast:Glyco_32</a>	Q8W431_WHEAT 84-559	88	563	0	
<a href="#">SCOP:d1gyha_</a>	d1gyha_	92	421	7e-62	▼

Click on a name to highlight the feature in the diagram above. Click the feature name for more information.

## Glyco\_32 domain

This is a SMART **Glyco\_32** domain ([full annotation](#)).

Glyco\_32

**Position:** 88 to 563

**E-value:** 5.6538458160159e-222 ([HMMER2](#))

**SMART ACC:** [SM000640](#)

**Definition:** Glycosyl hydrolases family 32

**Description:**

**Interpro abstract ([IPR001362](#)):** O-Glycosyl hydrolases ([EC 3.2.1.](#)) are a widespread group of enzymes that hydrolyse ...([full abstract](#))

**GO process:** carbohydrate metabolic process ([GO:0005975](#))

**GO function:** hydrolase activity, hydrolyzing O-glycosyl compounds ([GO:0004553](#))

### Glyco\_32 domain sequence (476 aa):

[Submit to BLAST](#)

[Align with the SMART alignment](#)

[Copy to clipboard](#)

```
HFQTAKNYMSDPNGLMYINGWYHMFQYNPVGTDWDDGMEWGHAVSRNLVTWRTLPIAMV
ADQWYDILGVLGSM TVLPNGTVIMIYTGATNASALEVQCIATPADPNDPFLRRWTKHPA
NPVIWSFPPGIGTKDFRDPMTAWYDESDDTWRTLLGSKDDHDGHHDGIAMMYKTKDFLNYE
LIPGILHRVQRTGEWECIDFYVVGHRSDNSSEMLHVLKASMDDERHDYISLGTYDSAAN
TWTPIDPELDLIGLRYDWGKPYASTSFYDPAKRRVLMGYVGEVDSKRADVVKGWASIR
SVPRTIALDEKTRTNLLLWPVEIETLRLNATELSDVTLNTGSVIHIPLRQGTQLDIEAT
FHLDASAVAALNEADVGYNCSSSGGAVNRGALGPPGLLVLAAGDRERGEQTAVYFYVSRGL
DGG LHTSFCQDELRSRAKDVTKRVIGSTVPLDGEAFSMRVLVDHSIVQGFAMGG
```



# Pfam预测蛋白活性位点

第98位的天冬氨酸为活性位点



天冬氨酸



## Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
<a href="#">Glyco_hydro_32N</a>	Glycosyl hydrolases family 32 N-terminal ...	Domain	<a href="#">CLO143</a>	88	407	88	407	1	308	308	345.6	2.5e-103	98	<input type="button" value="Show"/>
<a href="#">Glyco_hydro_32C</a>	Glycosyl hydrolases family 32 C terminal	Domain	n/a	451	563	477	563	6	86	86	39.4	5.7e-10	n/a	<input type="button" value="Show"/>

## Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length
				Start	End	Start	End	From	To	
<a href="#">Glyco_hydro_32N</a>	Glycosyl hydrolases family 32 N-terminal ...	Domain	<a href="#">CLO143</a>	88	407	88	407	1	308	308
#HMM	HfapekgwmNDPnglvyykgeYhlfYQynpkavwg nisWghavSkDlvhWeelpvalapdeeydknGvSGsavvlpd.klvllYtntkke.eqvqnlayskdkg...rewtkyeldknpvlvekpee.ntkdfERDPkvwake..dgkwrwvvgakdk...kkgkvllykskdlkewelasellhsseekaggnwEcpdlfpvsek.seeekkvkchv									
#MATCH	Hfgt+kt+m DPngltyygtYh+f+Qynp gt wt ++WghavSt+lv+W+lp+at++d+tydt Gv SGst+vlpt +++++YTg t+ + tvqt+at++dtt rtwklttt npvt + ptt +tkdERDP taw++ dt+wr++t+gkdt +tg++++Yk+kt+ +tel + +lh+ + +tg wEctd+tpv ++++++ hv									
#FP	*****889*****99*****877*****996.****7.5556*****967*****99888899*****98.7*****997778899****									
#SEQ	HPQTAKNYMSDFPGLMYHGWYHMPFQYNPVGTDWdGMEWGHAVSRNLVTRKLPFIAMVADQWYDILLVLSGSMVLPNGTVIMITYGATNASLLEWQCIATFADFNpFLRRWTKQPA-NPVIWS-PRGtGTRDFRDFMTAWYde=DDTWKTLVLSKDDKdghdGLAMMYTKDFLNYELTPGLHRVW-NIGEWFECLDFYVGHRSNDRSSEMLHV									
	-IDERHDIYSLGTVDAAANTWDFEpaal dLGLGLRYDWGR-PYASTSFYDFAKKRRVLMGYVGEVDSKGRADVV-KGASISRSVFRITLDELERTNLLWF									
<a href="#">Glyco_hydro_32C</a>	Glycosyl hydrolases family 32 C terminal	Domain	n/a	451	563	477	563	6	86	86
#HMM	s&gslgkvlas.gslseesttiyyd.....ksneslclDrsss..advfkekfgsfvesdrllfeekdkdlsrlilyDhssvEvFangg									
#MATCH	a g+gl+vla + +e+tt+tytt +tstc D+ rss +dv kt gsv++ + + + str+lvDhst+ Fa gg									
#FP	455788****9978899*****9*****9888888888888888.443...456*****99999887									
#SEQ	SALGFCLLVLAa@DRRGEQTAVYFYvsrgldgGLHHSFCQDELRSF&KDVTKRVIQSTVFPV-IDG---F&FSMRVLDHHSIVQG&FAMG									



# MEME

Multiple Em for Motif Elicitation

## MOTIF 1

[Next Top](#)

### Summary ?

E-value	6.8e-288
Width	50
Sites	9
Log Likelihood Ratio ?	1194
Information Content ?	182.6 (bits)
Relative Entropy ?	191.4 (bits)

### Sequence Logo ?



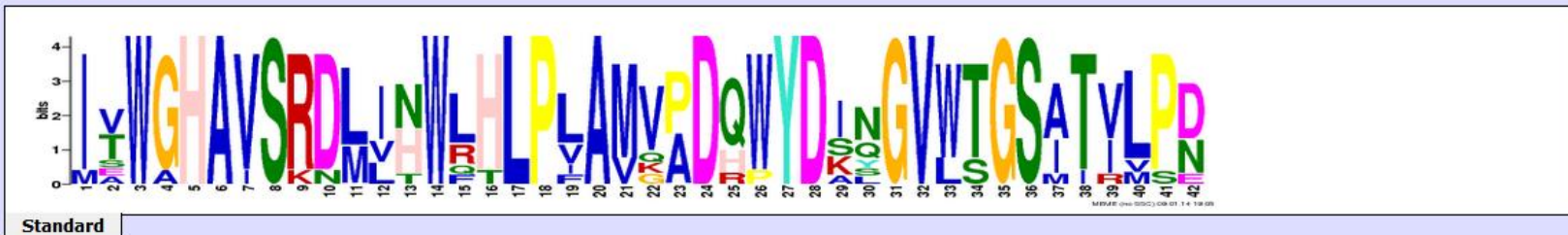
## MOTIF 2

[Previous](#) [Next Top](#)

### Summary ?

E-value	1.8e-192
Width	42
Sites	9
<a href="#">show more</a>	

### Sequence Logo ?



## MOTIF 3

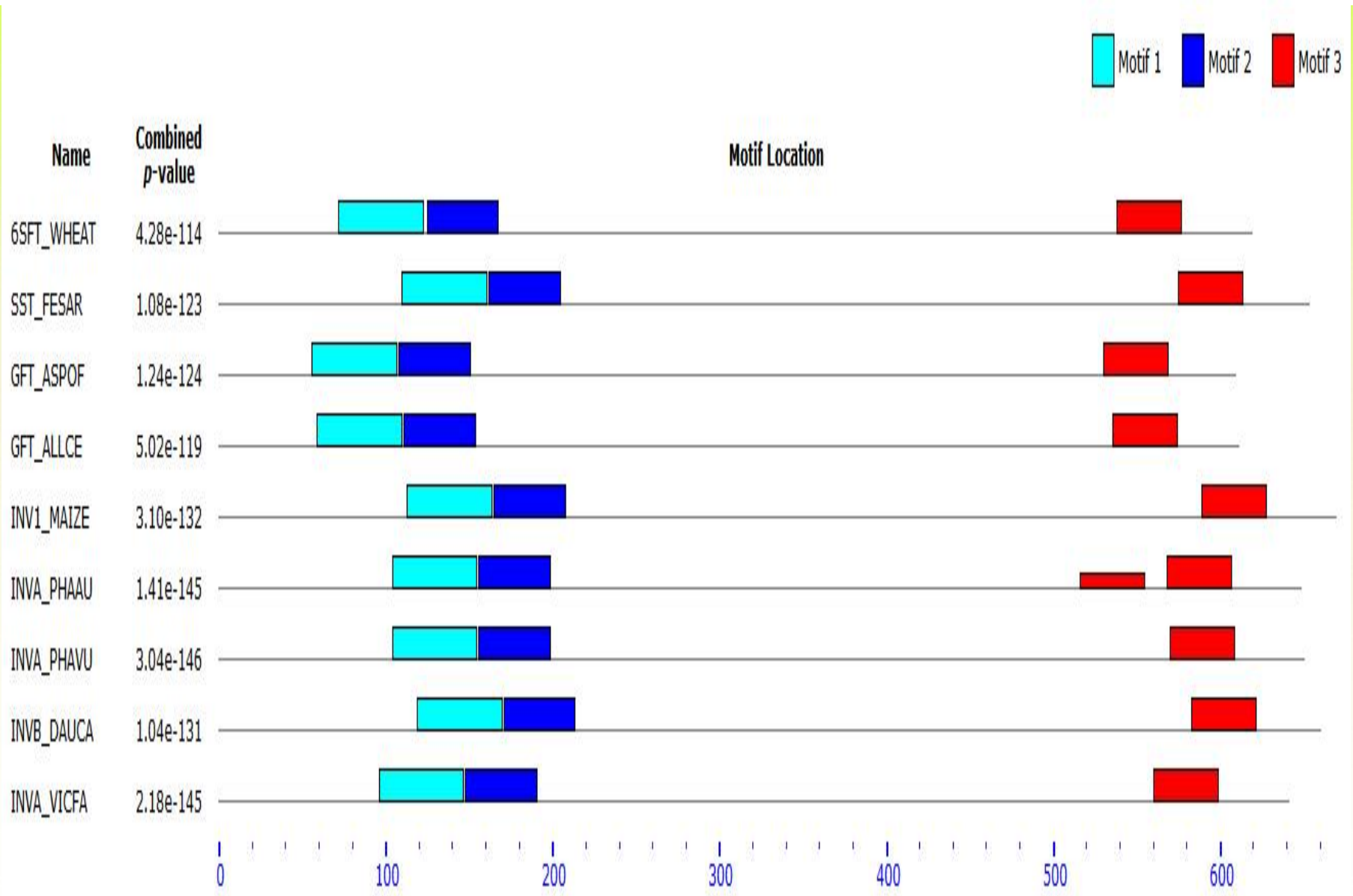
[Previous Top](#)

### Summary ?

E-value	1.1e-184
Width	38
Sites	9
<a href="#">show more</a>	

### Sequence Logo ?

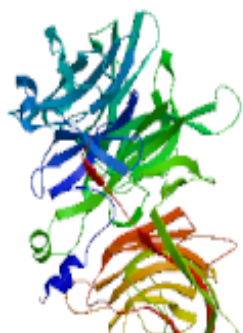




# SWISS MODEL预测三级结构



以3ugfB为模板,对小麦的6-SFT进行同源模建

## Model Summary



### Model information:

Modelled residue range: 72 to 612  
Based on template: [3ugfB] (1.70 Å)  
Sequence Identity [%]: 51.64  
Evalue: 0.00e-1

Quality information: [details]   
QMEAN Z-Score: -2.31 

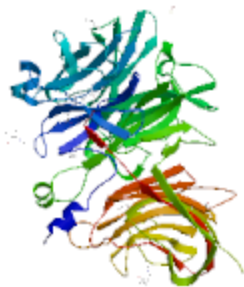
### Quaternary structure information: [details]

Template (3ugf): MONOMER  
Model built: SINGLE CHAIN

### Ligand information: [details]

Ligands in the template: GOL: 4, NAG: 5, SO4: 1.  
Ligands in the model: none.

## 3ugfB: [ Parent PDB: 3ugf Chain: B]



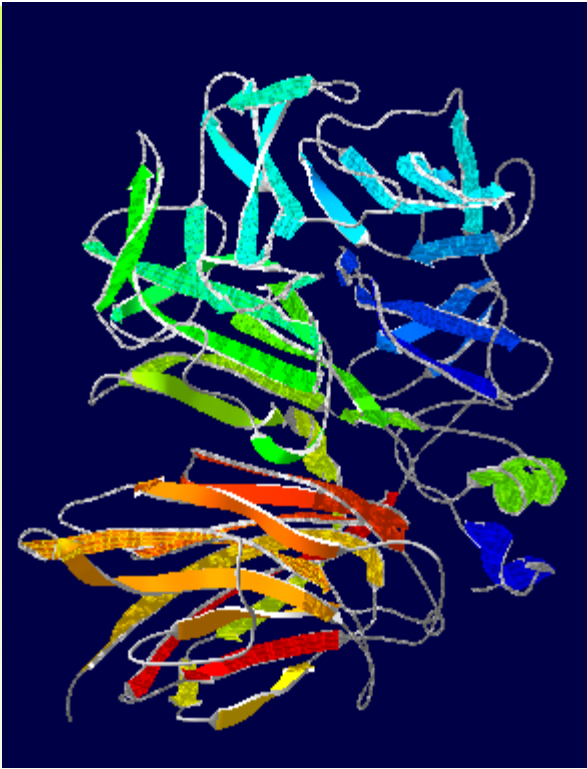
CRYSTAL STRUCTURE OF A 6-SST/6-SFT FROM PACHYSANDRA TERMIN  
Experiment: X-RAY Resolution: 1.70  
W.LAMMENS,A.RABIJNS,A.VAN LAERE,S.V.STRELKOV,W.VAN DEN END  
CRYSTAL STRUCTURE OF A 6-SST/6-SFT FROM PACHYSANDRA TERMIN  
Protein Quaternary Structure: NO-ENTRY

[ RCSB ] [ PDBsum ] [ PDB ] [ PISA ]

[ SCOP ] [ CATH ]

[ open with Swiss-Pdb-Viewer ] [ save to disk ]

# Swiss-pdbViewer

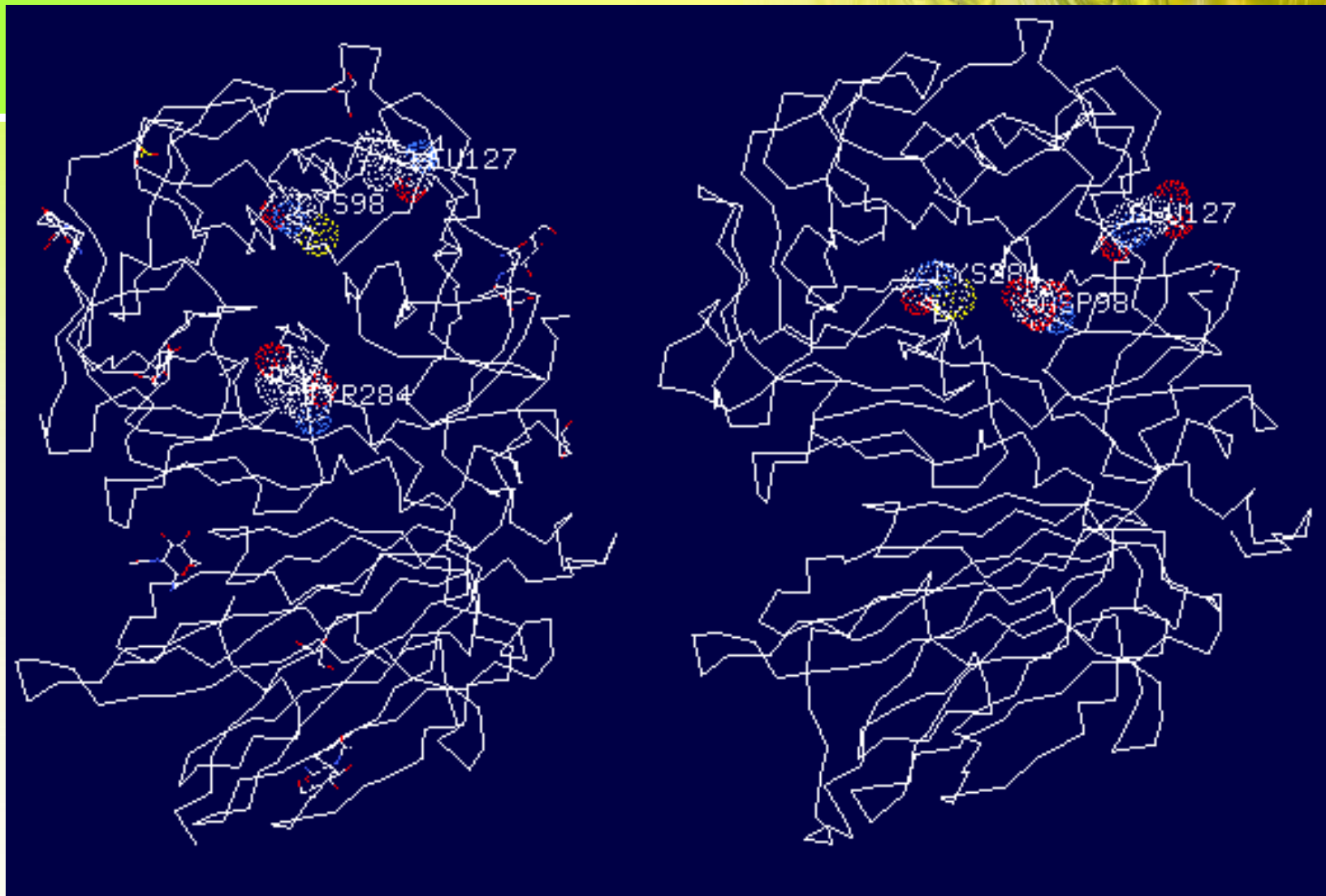


**3ugfB-pachysandra**



**6-SFT-WHEAT**

选取两个分子中 $\alpha$ 碳进行结构叠合均方根误差  
(RMSD) 为1.79埃



**3ugfB-pachysandra**

**6-SFT-WHEAT**

# 结论

## 6-SFT特点:

1. 该蛋白是一种亲水性蛋白质、只有一个跨膜螺旋、亚细胞定位在线粒体中;
2. 二级结构主要以无规则结构为主;
3. 结构域为糖基水解酶家族32;
4. 第98位的天冬氨酸为活性位点;

## 6-SFT功能:

1. 参与果聚糖合成, 催化蔗糖转成果糖和霉菌赤藓糖;
2. 有水解酶活性;

# 参考文献

- [1]张慧, 董伟, 周俊马. 果聚糖蔗糖转移酶基因的克隆及耐盐转基因烟草的培育. 生物工程学报. 1998, 2:181-186
- [2]殷桂香, 余茂云, 高翔. 植物果聚糖合成酶基因克隆及特性分析. 中国生物工程杂志. 2009, 29 (2) :125-133
- [3]岳爱琴, 李昂, 毛新国. 小麦果聚糖合成酶基因6-SFT-A单核苷酸多态性分析及其定位. 2011, 44 (11) :2216-2224
- [4]French AD, Waterhouse AJ. chemical structure and characteristics . In :Suzuki M, Chatterton Nj, eds. Science and technology of fructans .Florida:CRC Press 1993, 41-81



# 致谢

- 衷心感谢罗老师在学习上对我们的关怀和帮助,在此献上我们最崇高的敬意和最诚挚的谢意!
- 感谢秋季班同学整个学期的一路相伴!
- 感谢我的G11团队,生活因你们而精彩!