

# 褐飞虱 Bt 受体——碱性磷酸酯酶 (ALP) 的生物信息学分析

报告人：王增霞  
其他成员：张方梅  
徐秀秀  
陈雅琼

2013.1.11

# 报告内容

---

1

研究背景

2

研究目的与意义

3

报告内容

# 1

## 研究背景



**褐飞虱：同翅目，飞虱科。**

**别名：褐稻虱，稻飞虱的一种。  
褐飞虱为单食性害虫，只能在水稻和普通野生稻上取食和繁殖后代。**



**对水稻的为害主要表现在以下几方面：**

- ①直接吸食为害**
- ②产卵为害**
- ③传播或诱发水稻病害**

# 1

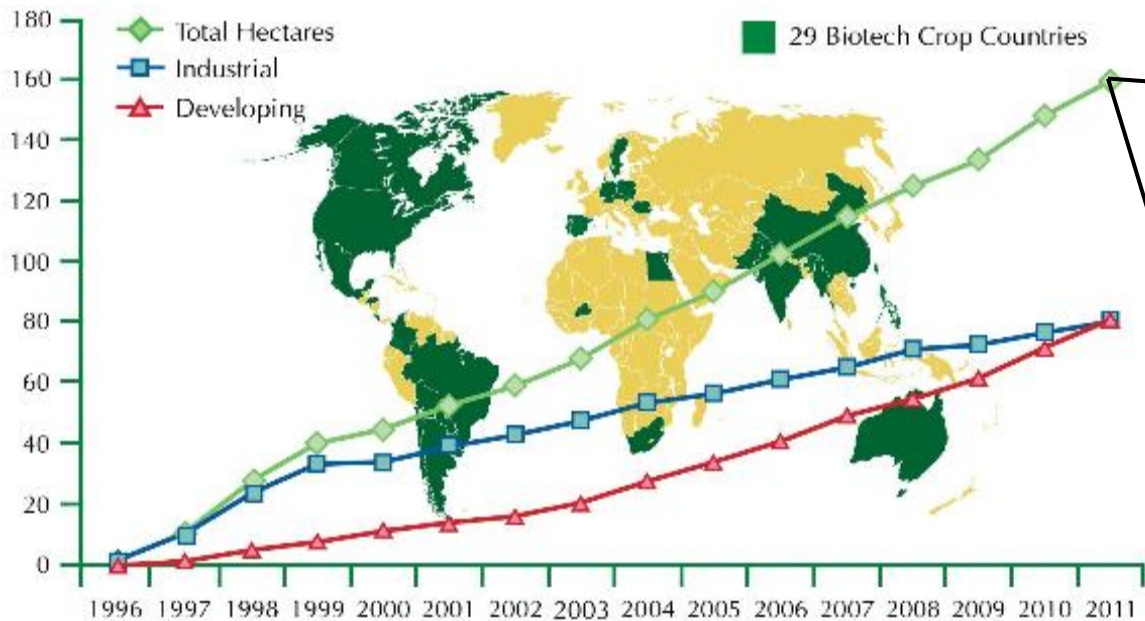
## 研究背景

- 稻飞虱是水稻上最主要的害虫之一，给水稻的生产造成了严重危害。

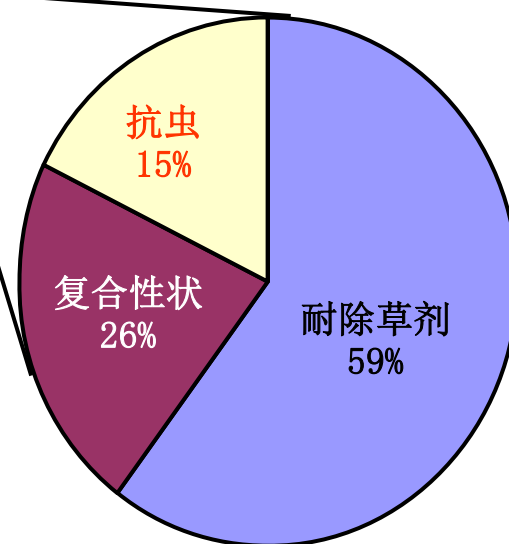


# 1

# 研究背景



2011年全球种植情况



A record 16.7 million farmers, in 29 countries, planted 160 million hectares (395 million acres) in 2011, a sustained increase of 8% or 12 million hectares (30 million acres) over 2010.

转基因作物在全球的种植面积（1996-2011年），单位：百万公顷  
(ISAAA, James, 2011)



# 研究背景

---

- **许多转基因水稻已经获准田间试验，在美国和伊朗转基因水稻已经开始商业化种植，2009年底中国政府为转基因抗虫水稻“华恢1号”和“Bt汕优63”颁发了安全证书，这些转基因水稻主要控制鳞翅目的二化螟、三化螟、稻纵卷叶螟等，对非靶标害虫稻飞虱的影响有待于研究。**

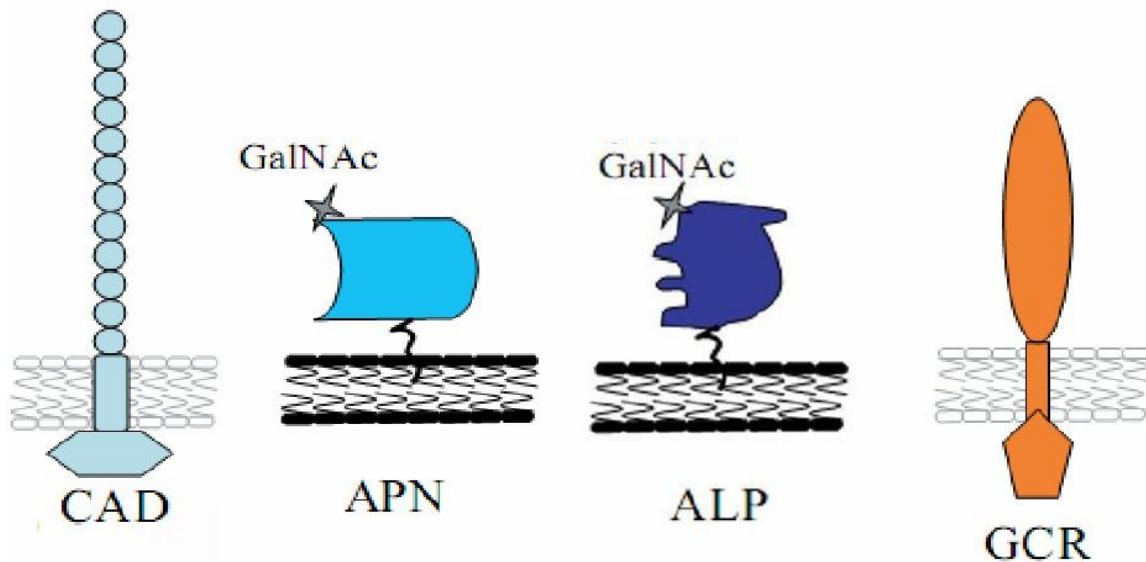
- Bt (*Bacillus thuringiensis*, 简称Bt) 产品以其高效安全、对目标害虫的特异性而倍受青睐，并逐渐成为研究最深入、应用最广泛的生物杀虫剂。



# 1

## 研究背景

### Bt受体蛋白



碱性磷酸酯酶同氨肽酶特性相似，也能够被糖基化的磷脂酰肌醇GPI锚定在细胞膜表面。目前已经在烟草天蛾、棉铃虫和家蚕等多种昆虫中证明碱性磷酸酯酶是Bt的受体之一。





- 目前关于Bt防治害虫的研究主要集中在鳞翅目、鞘翅目等防治效果显著的害虫中，带来了很好的经济和生态效益。
- 尚未见到对稻飞虱有毒力Bt菌株的报道，为实现利用Bt类产品防治稻飞虱，发展转Bt基因水稻，本研究从褐飞虱的Bt受体ALP的研究入手，利用一些生物信息学软件对稻飞虱的ALP进行分析，以揭示Bt毒素在防治稻飞虱方面的可能性。

## ALP的研究进展

- **ALP是一类特异性的磷酸水解酶，具有催化磷酸集团的转移反应和磷酸单酯酶水解反应的活性，且能被糖基化的磷脂酰肌醇GPI锚定在细胞膜的表面。**
- **Eguchi等（1995）报道在家蚕中存在两种ALPs，即膜结合的ALP、可溶性的ALP，其氨基酸同源性的60%。**  
**Itoh等（2003; 2008）也发现在家蚕和野蚕中存在ALP-m、ALP-s两种形式的ALP基因，并且这两个基因的长度不同。**

## ALP的研究进展

年代	作者	昆虫	结合Bt毒素
2003	McNall, Adang	烟草天蛾	Cry1Ac
2004	Jurat-Fuentes, Adang	烟蚜夜蛾	Cry1Ac
2006	Fernandez	埃及伊蚊	Cry11Aa
2009	Sarkar	棉铃虫	Cry1Ac
2010	Ning	棉铃虫	Cry1Ac
2011	Likitvivatanavong	埃及伊蚊	Cry11Ba

2

搜索褐飞虱的EST数据库，查找与家蚕ALP的氨基酸序列同源性在40-50%的2个同源片段，基于这两个片段，设计引物，获取两条ALP全长基因，分别记为ALP-1和ALP-2。

3

构建系统进化树



# 报告内容

## 引物设计

- 在有EST文库或者基因组文库的情况下：

搜索序列 → 设计特异性引物 → PCR扩增全长

- 在完全未知序列的情况下：

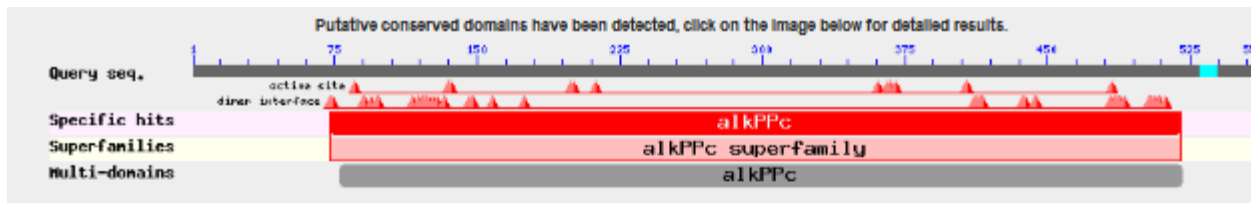
先搜索相关的其他物种的该序列 → 设计简并引物

→ PCR扩增全长

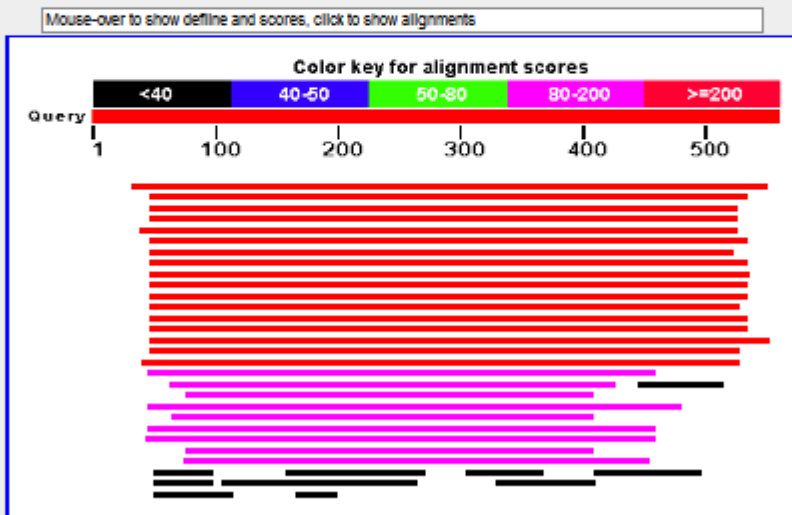
# 3

## 报告内容

### 利用Blast搜索ALP同源性基因



Distribution of 36 Blast Hits on the Query Sequence



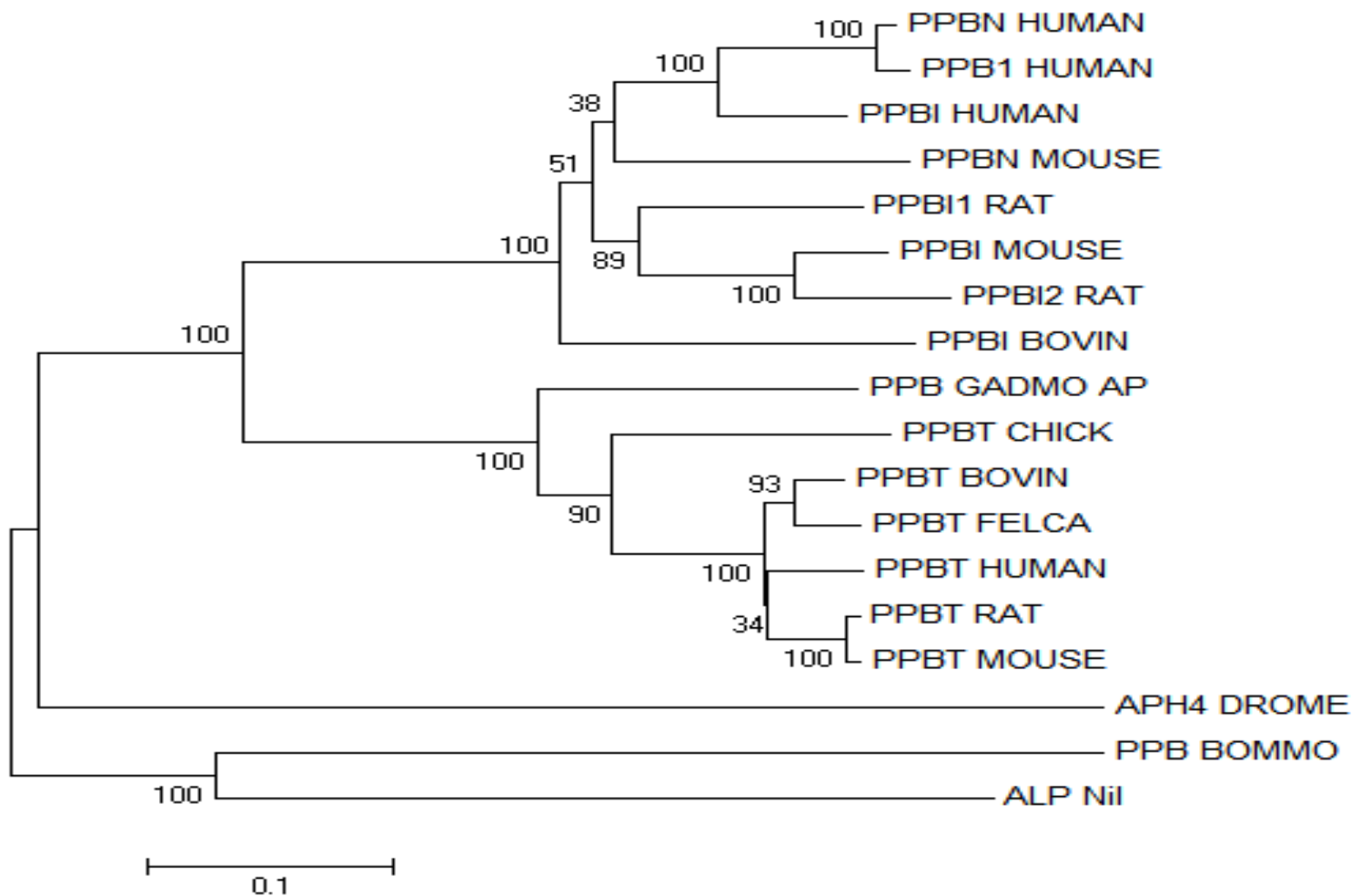
```

>PPB_BOMMO
MSTWWLVVVAAGLVRAEDRYHPERLAGEASAATRSAAESEASFV
VKVATDALKEDSLVVVADHHTVMSFNGYSRPGTDVLGTVRSLDSNRMP
>PPBT_CHICK
MKAFLLTLAQLCSASLVPEREKDPEYWRQQAQETLRDALRLQHLNQVAK
FGGYTPRGNPIFGLAPMQSDVDRKPFSTILYGNPGYKIVGGERENVSADF
>PPBT_BOVIN
MISPFLLAIGTCFASLVPEKEKDPKYWRDQAQQLTKNALRLQTLNTVAKN
FTFGGYTPRGNISIFGLAPMVSDDTKKPFSTAILYGNPGYKVVGGERENSMV
>PPBT_RAT
MILPFLVLAIGTCLTNSFVPEKEKDPKYWRQQAQETLKNALQKLNQVAKN
GGYTPRGNISIFGLAPMVSDDTKKPFSTAILYGNPGYKVVGGERENSMVDY
>PPBT_MOUSEMISPFVLAIGTCLTNSFVPEKERDPSYWRQQAQETLKNAL
TVVADHSHVFTFGGYTPRGNISIFGLAPMVSDDTKKPFSTAILYGNPGYKVV
>PPBT_FELCA
MISPFVLAIGTCLTNSLVPEKEKDPKYWRDQAQQLTKNALRLQKLNQVAKN
TFGGYTPRGNISIFGLAPMVSDDTKKPFSTAILYGNPGYKVVGGERENSMVD
>PPB_GADMO
APAGFPEQEPEPKFVNDWAQKTLDKALSQTLNKNKAQNLILFLGDGMGV
AAFISDVQKPFSTAILYGNPGYKLVNGARENSTVDYQDNSYLAQAAVPLS
>PPBT_HUMAN
MISPFVLAIGTCLTNSLVPEKEKDPKYWRDQAQETLKYALELQKLNQVAKN
FGGYTPRGNISIFGLAPMLSDTKKPFSTAILYGNPGYKVVGGERENSMVDY
>PPBI_BOVIN
MQGACVLLLLLGLHLQLSLGLVPVEEEDPAFWNRQAAQALDVAKKLQPIQTA
ADHSHVFSFGGYTLRGTSIFGLAPSKALDSKYSYLSILYGNPGYALGGGSRPDV
>PPBN_HUMAN
MQGPWVLLLLLGLRLQLSLGIIPVEEENPDFWNRQAAEALGAACKLQPAQTA
GGYPLRGSSIFGLAPGKARDRKAYTVLLYGNPGYVLDGARPDTVTESESGS
>PPBI_HUMANMQGPWVLLLLLGLRLQLSLGVIPVEEENPAFWNRQAAEAL
TLTLVTADHSHVFSFGGYTLRGSSIFGLAPSKAQDSKAYTSILYGNPGYVFN
>PPBI1_RATMQGDWVLLLLLGLRIHLSFGVIPVEEENPVFVWQKAKEALDV
ITADHSHVFAFGGYTLRGTSIFGLAPLNAQDGKYSYLSILYGNPGYVLSNGNR
>PPB1_HUMAN
MLGPCMLLLLLLGLRLQLSLGIIPVEEENPDFWNRQAAEALGAACKLQPAQTA
SFGGYPLRGSSIFGLAPGKARDRKAYTVLLYGNPGYVLDGARPDTVTESESG
>PPBI_MOUSE
MQGPWVLLLLLGLRLQLSLVIPVEEENPAFWNKKAAEALDAACKLQPIQTS
YTLRGTSIFGLAPLNALDGKPYTSILYGNPGYVGTGERPNVTAESSGSSYR
>PPBN_MOUSE
MWGACLLLLLGLSLQVCPVIVVEEENPAFWNRKAAEALDAACKLQPIQTS
HVFSGGYTQRGASIFGLAPFKAEDGKSFTSILYGNPGYKLVHNGARADVTEE
>PPBI2_RATMQGAWVLLLLLGLFRLQLSLVIPVEEENPAFWTQKAAADALNV
DHSVFSFGGYTLRGTSIFGLAPLNALDGKPYTSILYGNPGYVGTGERPNVT
>APH4_DROME
MHCLVILGFLGSLVAFSWAGVTTQPPPLIRLTSAGGDTGPGQFDVGKTKEPED
HELVEFDLAIQAANNNDPDELTLVADHSHAVTFNGYALRGADILGTANS
  
```

# 3

## 报告内容

### 构建系统进化树



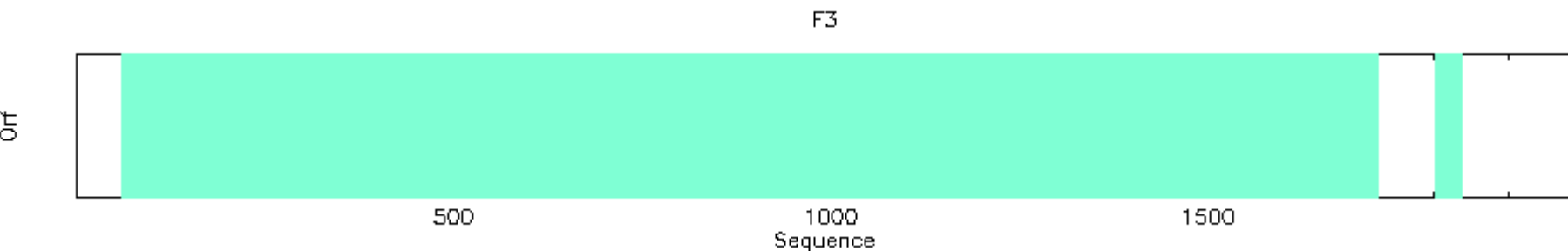
### 3

## 报告内容

### 4

## 核苷酸序列和蛋白质序列分析。

- 利用plotorf(v6.0.1) (Plot potential open reading frames in a nucleotide sequence)程序预测读码框，寻找潜在的ORF序列。



F3的ORF最长，可推测出ORF序列的长度大于1500bp





## 报告内容

---

- 利用ShowORF程序分析ALP全长mRNA序列读码框，可以得到1-61为5'UTR，62-1726为开放阅读框，1727-1780为3'UTR。

- 利用GetORF程序提取ALP全  
区核苷酸序列和所编码的氨基酸序列

## Additional section

Code to use:	Standard
Minimum nucleotide size of ORF to report:	1000
Maximum nucleotide size of ORF to report:	1000000
Type of output:	Nucleic sequences between STA

FVLSCSAVVRAMLAFSPLIIVAVLLSSLTLLQQLPETRCTPLGETALNPEINPYHYHSIDK  
 AVDLPYKDENEKYWTIQNAISSLQEKLEKSIKKKAKNLLFLGDGMSVPTVAAAARIYLG  
 QLEMKAGENSRLSFEFPFTGYSKTYCVDKQVADSACTSTAYIGGVKANYGTLGVTAAVA  
 RGNCTASLDPAHVIGVVTWAQTAGKATGIVTTASVTDASPAGAYSHSADRDWQSDKDLR  
 KAGAQAADACKDIAQQLVRNKPSELKVILGGGRKKFLPETVKDKSGVKGERLDGANLIQE  
 WLDDKKERKATAKYFEDRTIGLLESNTISSDYIMGLFRPEHMNYYLEPAAAGQPTLEEMIR  
 VAIERLQREPNGYLFEVAGALIDKAHQNWAQRALAEVFEFAKAIKLAESMTNKEDTILIV  
 VTADHAHTMSFSGYPRRGNILSTTPDHMKYSYSTLGYMNGPSAKSFQDAQNHFFENDH  
 MMQHDYRFPSTLQLEDETHGGDDVAVYARGPWSHLFVGTYEQHLLIPHMMFAAKIGAAGA  
 VGSSAIFTPSAVLLTISTLLLSFRIL

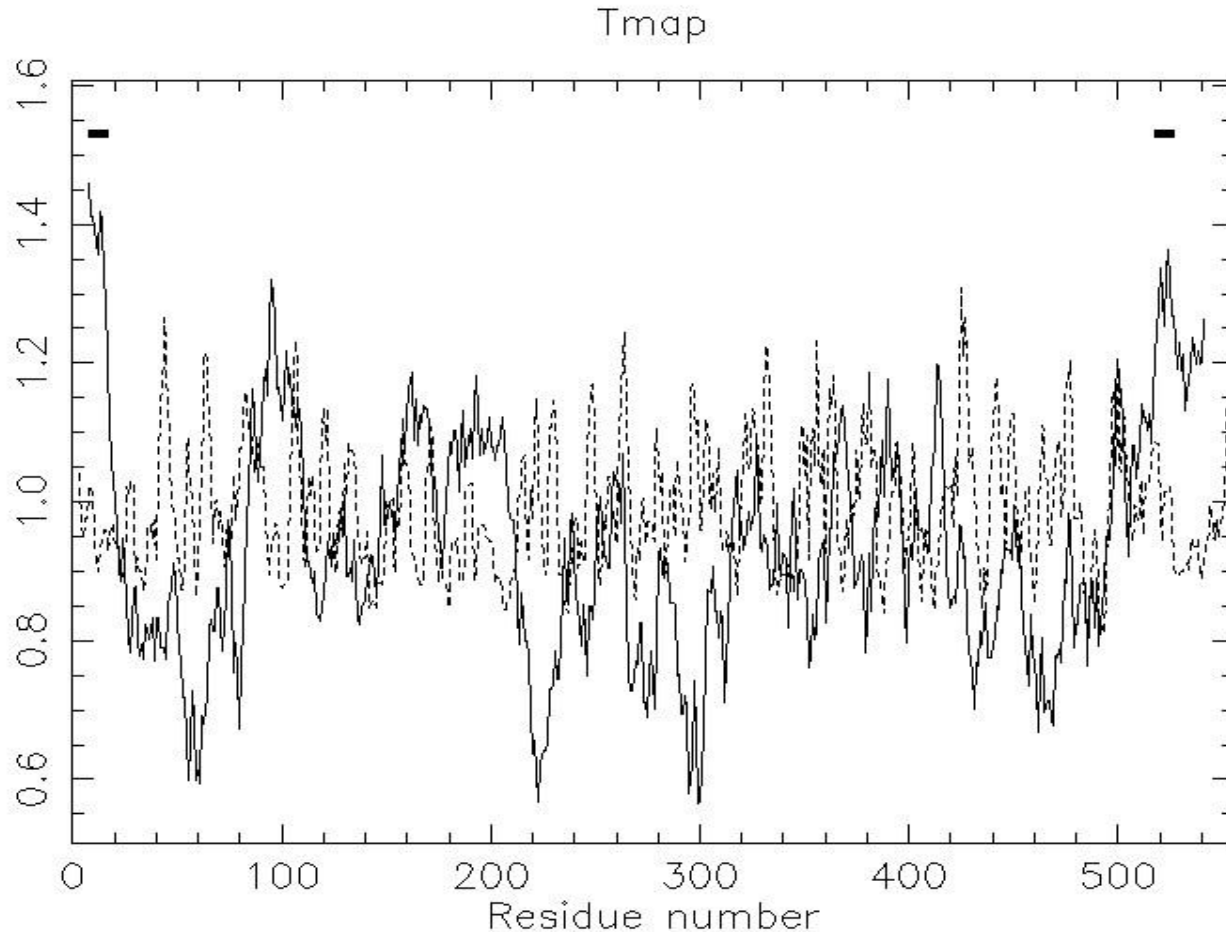
TTCTGTCAGCTGTTACAGTGTAGTCCGTGCAATGCTCGCCTTCAGCCCCTTATAATC  
 GCTGTGCGCCTTCTATCTTCCITGACACTGCTTCAACAATTGCCGAAACTAGATGTACT  
 CCACTGGGTGAAACCOCCTGAATCCCGAGAACCCTACTATCATTACTCAAITGATAAA  
 GCTGTAGATCTACCCTACAAAGATGAGAATGAATCAAAGTATTGGACCCAGAACGCAATC  
 TCATCCTTGCAGGAGAACTTGAGAAAAGTATTATAAAAAAGAAAGCGAAAAACCTAATC  
 CTATTTTTGGGAGATGGCATGTCAGTTCCTACAGTGGCAGCCGCGAGGATCTACTTGGGA  
 CAGCTTIGAGATGAAGGCTGGCGAGAATCCAGGCTGAGCTTTGAAGAGTTCCTTTCACT  
 GGCTATTCAAAGACATACTGCGTCGACAAACAGGTAGCCGATTTCGGCTGCACATCTACA  
 GCCTACACAGGCGGCGIGAAGGCCAACTATGGCACCTAGGCGTCACGGCGCCGTCGCC  
 CGTGGCAACTGCACAGCCAGCCTTACCAGCCACACACGTGACGGCGTCGTGACGTGG  
 GCGCAGACGGCCGGCAAGGCGACCGGCATCGTACGACGGCGTCGGTACCCGACGCGTCC  
 CCGGCCGGCGCATACTCGCACTCGGCCGACCGCGACTGGCAGAGCGACAAGGATTTGCGC  
 AAGGCTGGTGCAGGCGGACGCGTGC AAAGATATTGCGCAGCAGCTGGTGCACAACAAA  
 CCGGGATCTGAGCTCAAGGTIATCCTTGGTGGAGGAAGAAAAAAGTTCITGCCGGAAACA  
 GTGAAGGATAAATCTGGGGTGAAGGCGAGCGATTGGATGGAGCTAATTTGATCCAGGAG  
 TGGTTGGATGACAAGAAAAGAAAAGCTACTGCGAAATATTTGAAGACAGAACTGGA  
 CTCTAGAAATCAAATACTACTTCACTCCGACTATATAATGGGTTTATTCCTGCTGAGCAC  
 CTATCTGGAGCCTGCCGAGCAGGTCAGCCAACTTTGGAGGAGATGACTCGA  
 CGAAAGACTTCAACGAGAACC AAACGGTTACTACCTATTTGTTGAAGGAGCT  
 CAAAGCTCATCACCAGAACTGGGCTCAGCGCGCTCTAGCGGAGACTGTCGAG  
 AGCAATCAAGCTGGCAGAATCAATGACTAATAAGGAAGACACCCTGATAGT  
 AGATCAGCTCATACCATGTCCTTCAGTGGATATCCTCGTCGAGGGAATAAC  
 CACAACCCCTGACCACGACATGAAGTATAGCTACTCGACACTCGGCTACATG  
 CAGTGTAAAAAGTTTCCAAGACGCACAAAATCATATTTTAAAAATGACAAT  
 ACACGACTACAGGTTTCCTCTCTGTTCCAACCTGGAGGACGAGACGCACGGG  
 TGTGGCCGIGTACGCCAGAGGTCCTTGGTTCGATCTATTCTGTTGGAACTTAC  
 CCTCATCCCCACATGATGGCTTTTCCCGCCAAAATGGAGCAGCTGGAGCC  
 CTCGCCACTTTACCCCGTCCGCTGTCCTTCTTACAATCTCCACCTTGTCT  
 TAGAATATTG



## 报告内容

---

- 利用pepstats (v6.0.1) (Calculates statistics of protein properties)的功能是统计蛋白质序列的特征。  
蛋白质的分子量为60.92kDa,  
氨基酸的残基数555,  
等电点为6.5136,  
氨基酸的平均分子量为109.8,  
各个氨基酸的数目以及所占的摩尔百分比,  
氨基酸的酸碱性等。

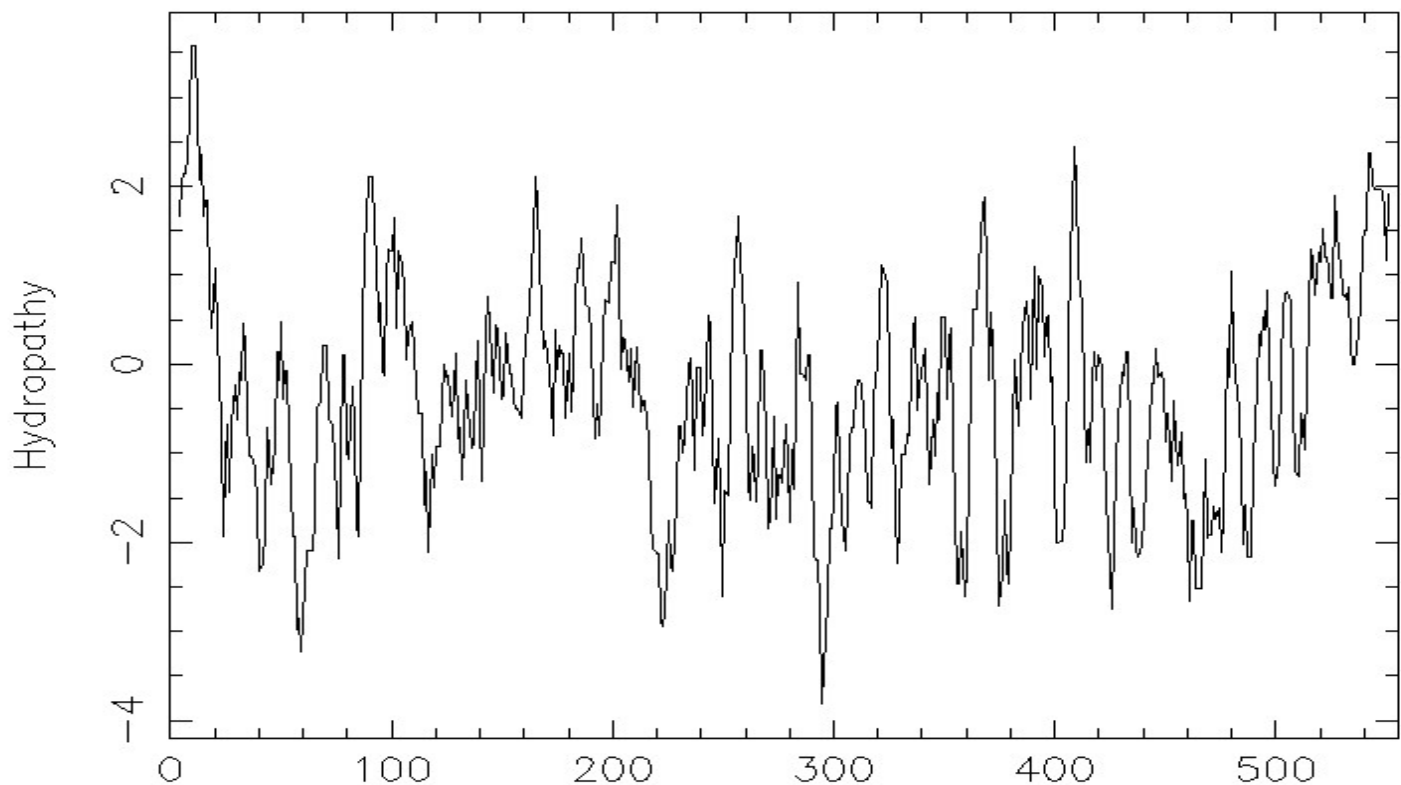


利用tmap程序分析蛋白质的跨膜结构，预测到2个跨膜的区域，都有18个氨基酸序列，分别为：

**4-21，跨膜序列为：FSPLIIAVCLLSSLTLLQ；**

**514-531，跨膜序列为：IPHMMAFAAKIGAAGAVG**

Kyte-Doolittle Plot

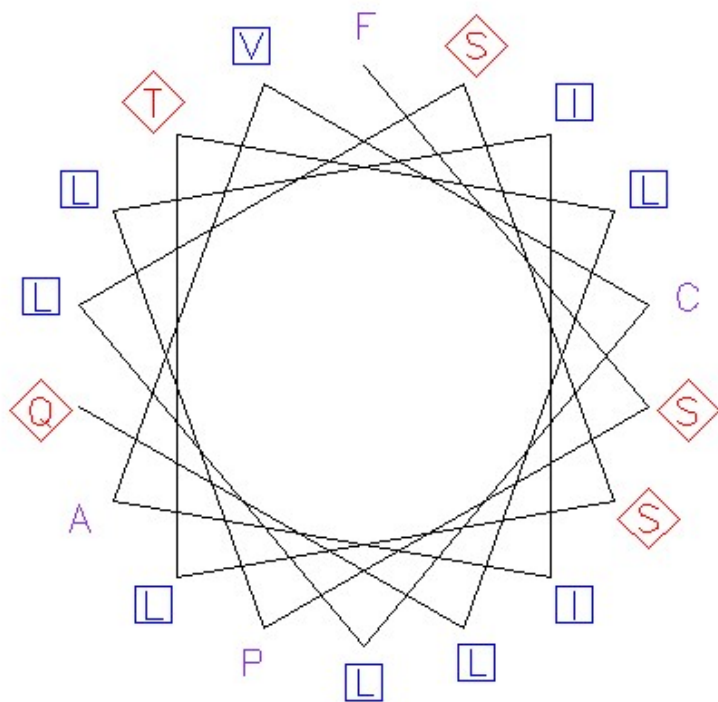


总体来看，该蛋白中疏水性氨基酸占的比例比较大，该预测结果与后面通过实验验证该蛋白是膜结合的**ALP**的事实相符，由于生物膜的有效成分是脂类，所以跨膜蛋白也应该是亲脂疏水的。

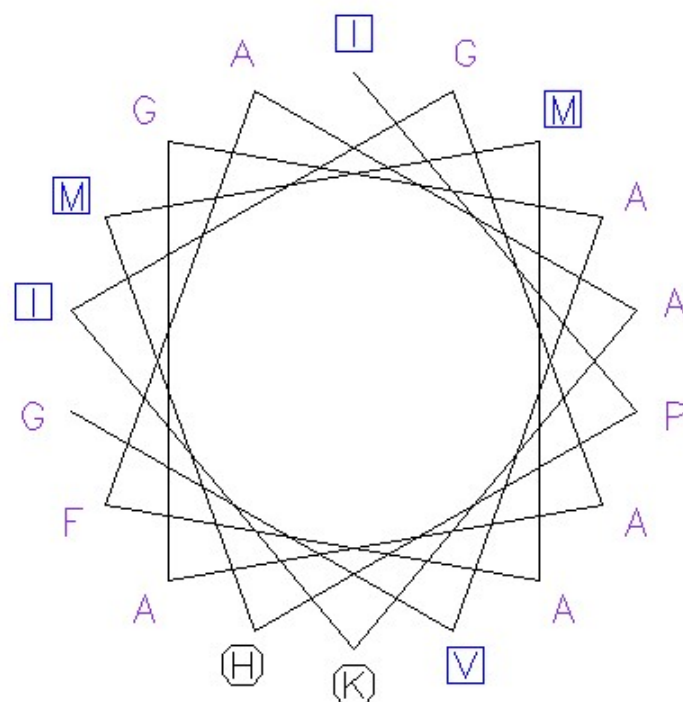
# 3

## 报告内容

- 利用pepwheel(v6.0.1), 可以分析两个跨膜区序列:



4-21跨膜序列 FSPLIIAVCLLSSLTLLQ

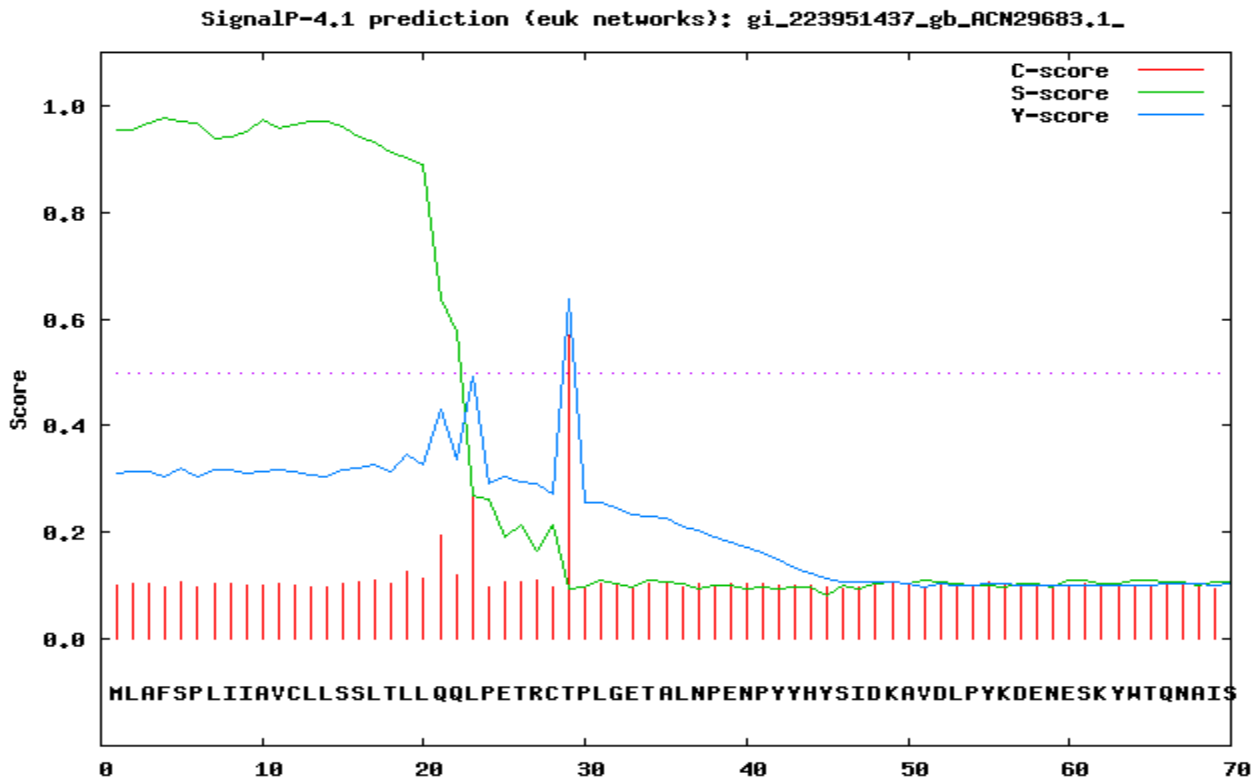


514-531跨膜序列 IPHMMAFAAKIGAAGAVG

# 3

## 报告内容

- 利用CBS的SignalP，可以信号肽的位置及强度：



#	Measure	Position	Value	Cutoff	signal peptide?
	max. C	29	0.569		
	max. Y	29	0.638		
	max. S	4	0.977		
	mean S	1-28	0.769		
	D	1-28	0.709	0.450	YES

Name=gi\_223951437\_gb\_ACN29683.1\_ SP='YES' Cleavage



# 报告内容

- 预测GPI锚定位点，发现该序列中包含一个GPI锚定位点。

[http://mendel.imp.ac.at/sat/gpi/gpi\\_server.html](http://mendel.imp.ac.at/sat/gpi/gpi_server.html)

Query sequence gi|223951437|gb|ACN29683.1| (length 555 amino acids): MLAFSPLIIA VCLLSSLTLL  
QQLPETRCTP LGETALNPEN PYYHYSIDKA VDLPYKDENE SKYWTQNAIS SLQEKLEKSI IKKKAKNLIL FLGDGMSVPT  
VAAARIYLGQ LEMKAGENSR LSFEEFPFTG YSKTYCVDKQ VADSACTSTA YTGGVKANYG TLGVTA AVAR GNCTASLDPA  
THVTGVVTWA QTAGKATGIV TTASVTDASP AGAYSHSADR DWQSDKDLRK AGAQADACKD IAQQLVRNKP GSELKVILGG  
GRKKFLPETV KDKSGVKGER LDGANLIQEW LDDKKERKAT AKYFEDRTGL LESNTTSSDY IMGLFRPEHM NYYLEPAAAG  
QPTLEEMTRV AIERLQREPN GYYLFVEGAL IDKAHHQNWA QRALAETVEF AKAIKLAESM TNKEDTLIVV TADHAHTMSF  
SGYPRRGNNI LSTTPDHDMK YSYSTLGYMN GPSAKSFQDA QNHFFENDNM MQHDYRFPSL FQLEDETHGG  
DDVAVYARGP WSHLFVGTYE QHLIPHMMAF AAKIGAAGAV GSSATFTPSA VLLTISTLLL SFRIL

```
Potential GPI-modification site was found.  
Quality of the site ..... : P  
Sequence position of the omega-site : 532  
Score of the best site ..... : 13.18 (PValue = 1.099638e-04)
```





谢谢!