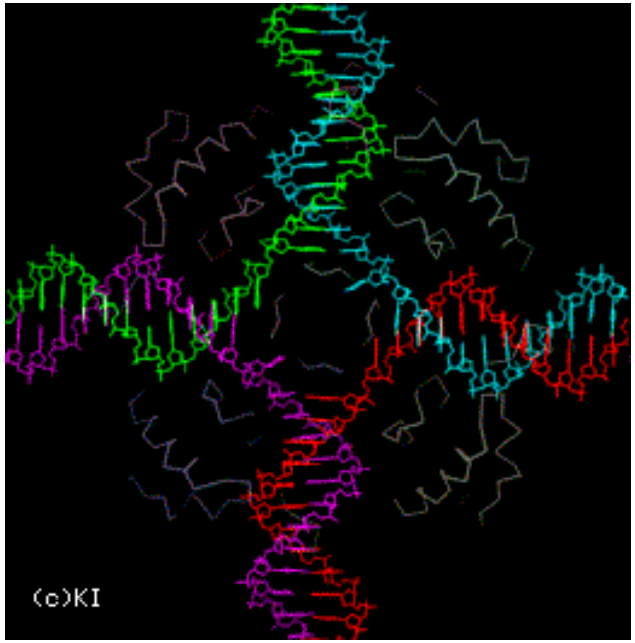
激活蛋白PEAt1的生物信息学分析



尹富仕 张丽勋 王垚 张薇
CAAS08S3 group2

中国农业科学院植保所药物工程室

激活蛋白（Activator）

- 该蛋白是从多种不同种属的植物病原真菌中筛选、分离、纯化出的一系列蛋白激发子（邱德文，2002）。
- 该类蛋白有较强的诱导植物抗病，增产，抗逆，促进植物生长的生物活性，提高作物产量，理化特性上该类蛋白都有很好的热稳定性。

几类蛋白激发子对比情况

蛋白激发子类型	主要生物来源	专利情况	分子量	过敏反应（烟草寄主）	热稳定性
隐地蛋白（Cryptogein）	卵菌（*P. cryptogea*）	法国专利1985年	10~15kD	或有或无	无报道
过敏蛋白（Harpin）	细菌（*E. amylovora*）	美国专利1992年	30~40kD	有	高
激活蛋白（Activitor）	真菌（*A. alternaria*）	中国专利2000年	30~70kD	无	很高

激活蛋白的作用机理

- 激活蛋白通过激活植物体内的免疫系统而达到抗病防虫和抗逆的目的。
- 激活蛋白通过激活植物的代谢系统而使植物生长加快，活力增强。脯氨酸含量的增加和纤维素酶的加强，植物细胞增长，根系的活力加强，达到增产目的。

激活蛋白（Activator）

- 激活蛋白可湿性粉剂已经商业化生产，大田推广中取得了良好的效果，具有广泛的应用前景。激活蛋白诱导处理种子能加快黄瓜种子的萌发，促进幼根的生长；提高多种瓜果蔬菜的座果率，如处理西瓜和番茄可使瓜果数和果重显著提高，分别达到77.4%和45.7%；，对白菜有显著的抗病增产作用，同时能大幅度提高白菜的品质（邱德文等，2005），同时，研究还发现该蛋白制剂能够成倍地提高Bt对鳞翅目害虫的防治效果，增效达18.5倍(邱德文，曾凡荣等，2005)。

研究背景

Peat1蛋白是一种从极细链格孢菌中分离得到的蛋白激发子（protein elicitor from Alternaria tenuissima，PEAt1）。经验证，该蛋白具有耐热的特性，能诱导植物产生系统抗性，促进植物生长，改善作物品质，可发展为新型的多功能生物农药。目前，该蛋白已申报中国发明专利，其有关基因已在GenBank登录（CH445335）。

分析目的

- 用生物信息学的方法分析PEAt1的结构
- 探讨PEAt1结构与功能的关系
- 分析与探究PEAt1耐热，抗病，促生长的分子机制
- 以PEAt1蛋白为例，练习和巩固有关生物信息学数据库和软件的使用，希望对大家分析自己研究中遇到类似的实际问题有所启发。

研究路线

1. 氨基酸的理化性质初步分析

2.膜蛋白的跨膜区预测

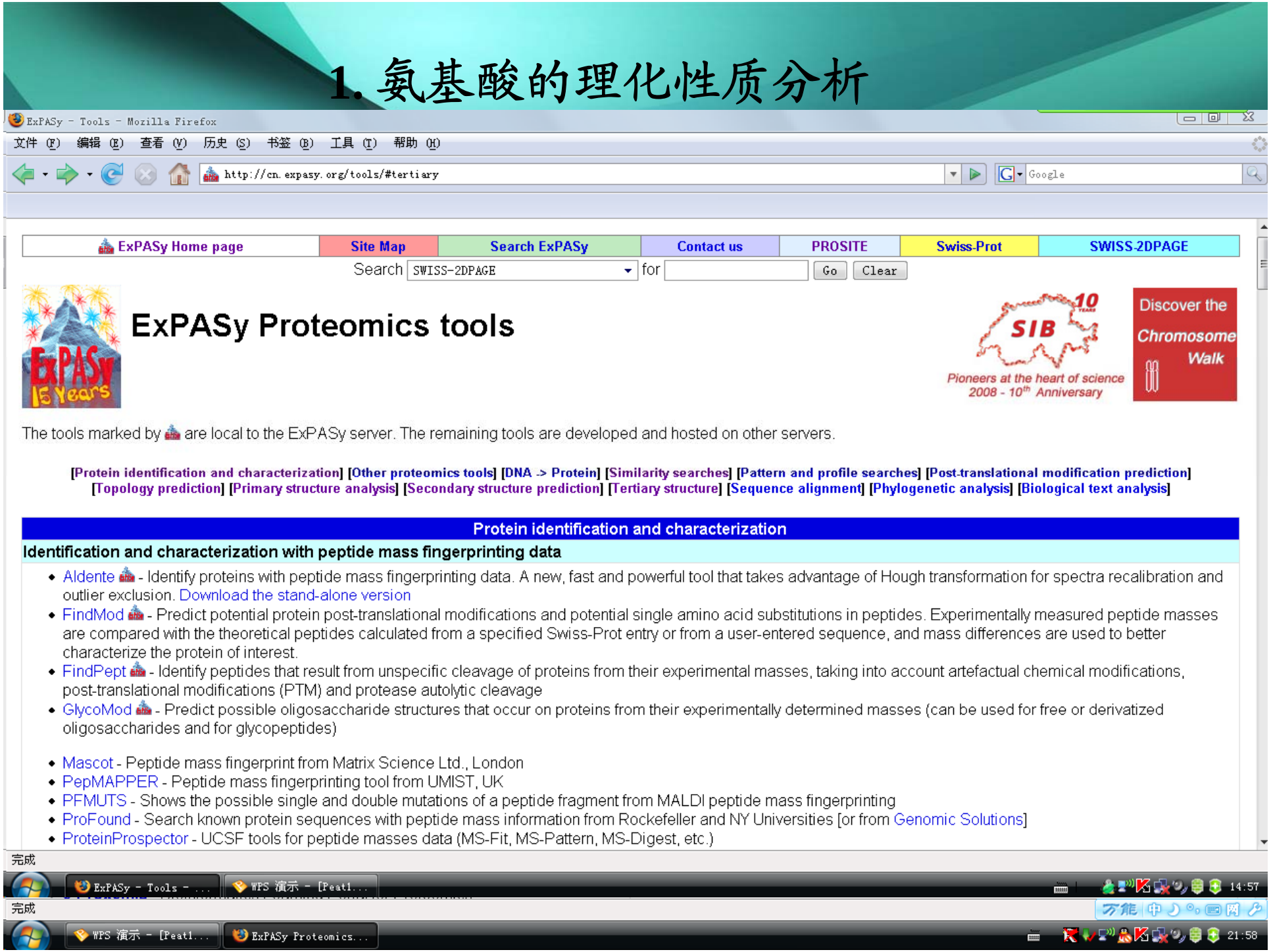
3.蛋白质序列的二级结构

4.蛋白质序列的结构域预测

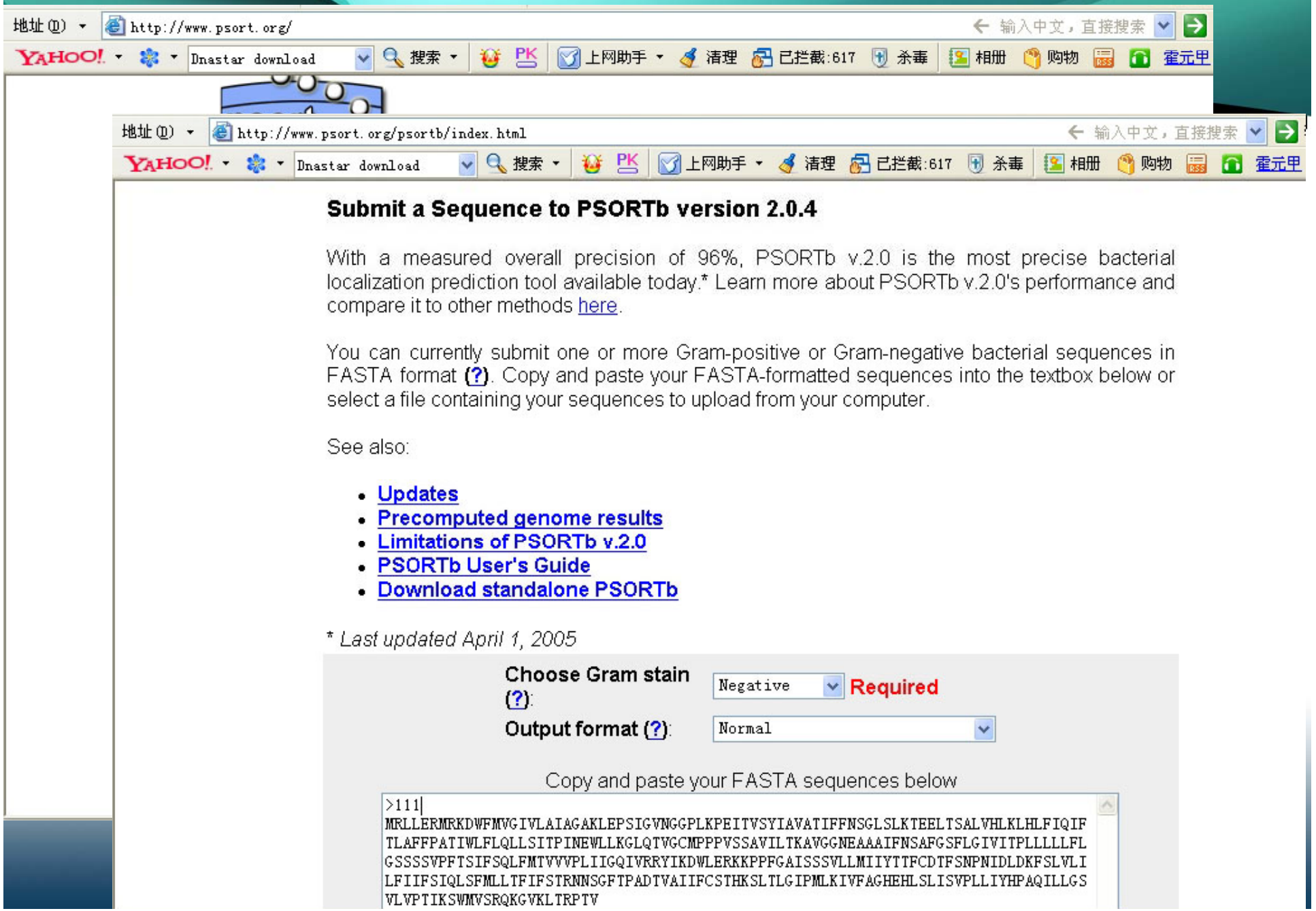
5.蛋白质序列的三级结构预测/同源建模

6.总结与讨论PEAt1结构与功能的关系

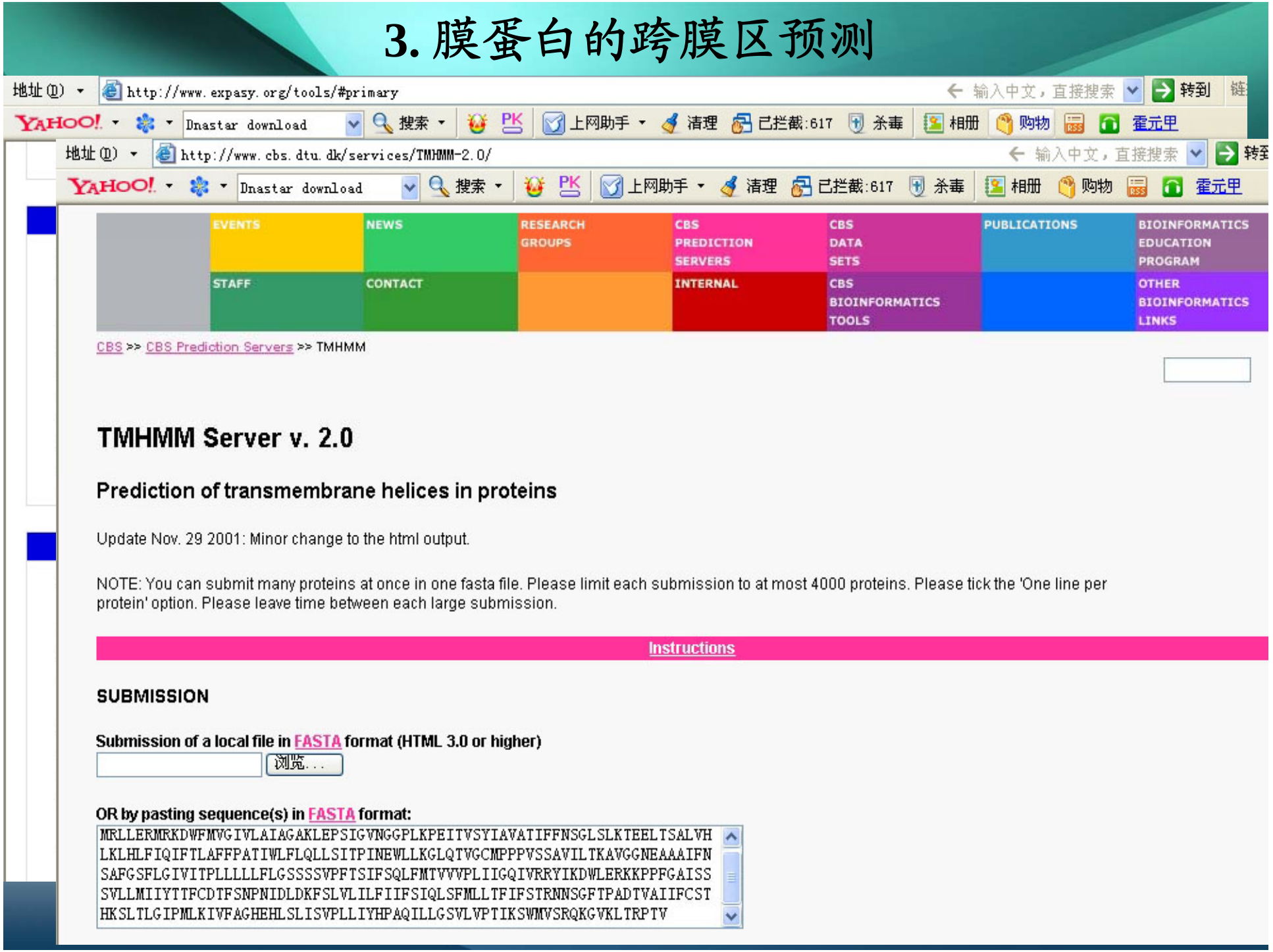
1. 氨基酸的理化性质分析



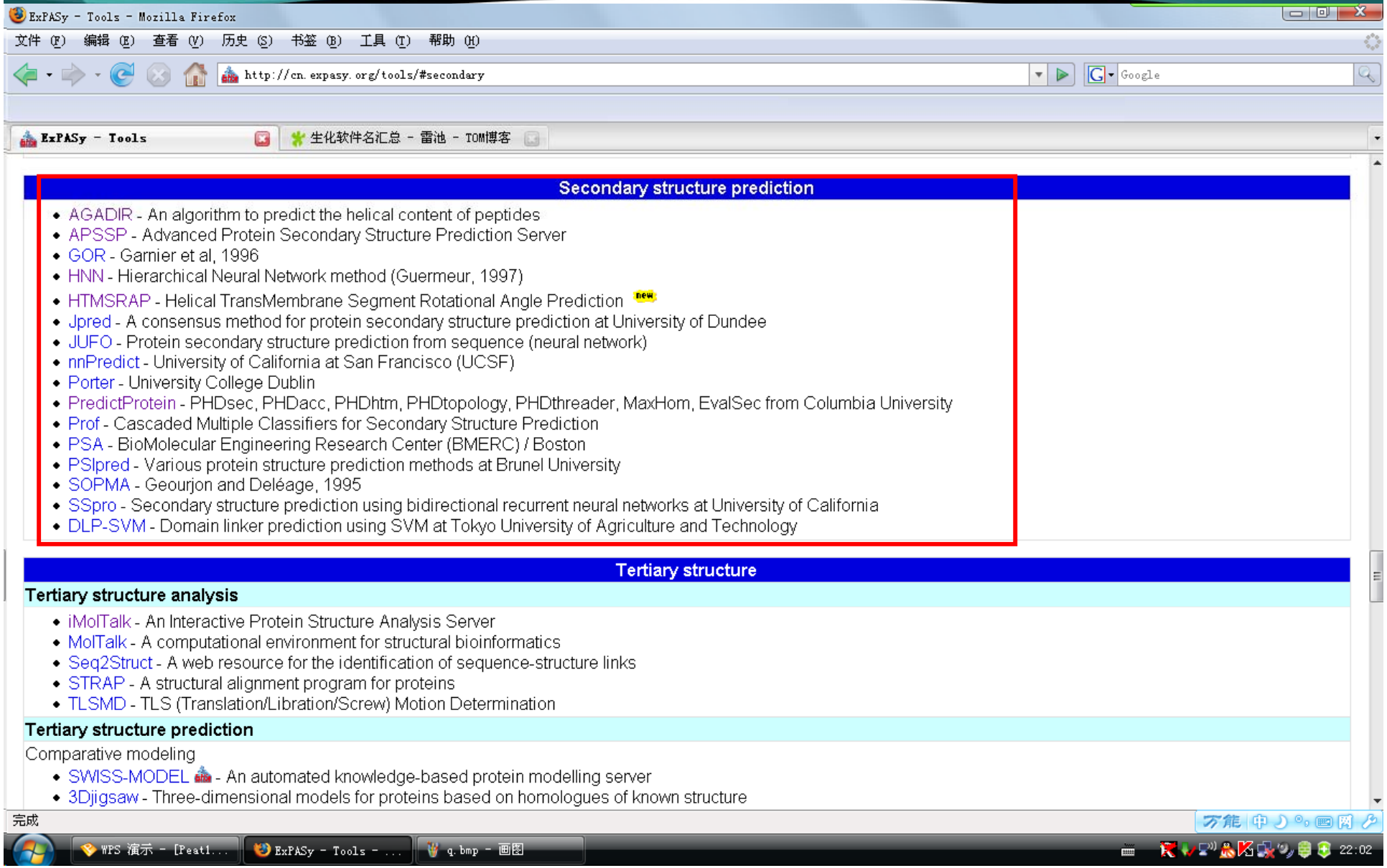
2. 蛋白质的亚细胞定位


地址(D) http://www.psort.org/
输入中文，直接搜索
YAHOO! Dnastar download 搜索 PK 上网助手 清理 已拦截:617 杀毒 相册 购物 霍元甲
地址(D) http://www.psort.org/psortb/index.html
输入中文，直接搜索
YAHOO! Dnastar download 搜索 PK 上网助手 清理 已拦截:617 杀毒 相册 购物 霍元甲
Submit a Sequence to PSORTb version 2.0.4
With a measured overall precision of 96%, PSORTb v.2.0 is the most precise bacterial localization prediction tool available today.* Learn more about PSORTb v.2.0's performance and compare it to other methods here.
You can currently submit one or more Gram-positive or Gram-negative bacterial sequences in FASTA format (?). Copy and paste your FASTA-formatted sequences into the textbox below or select a file containing your sequences to upload from your computer.
See also:
Updates
Precomputed genome results
Limitations of PSORTb v.2.0
PSORTb User's Guide
Download standalone PSORTb
* Last updated April 1, 2005
Choose Gram stain (?): Negative Required
Output format (?): Normal
Copy and paste your FASTA sequences below
>111
MRLLERMRKDWFMVGIVLAIAGAKLEPSIGVNGGPLKPEITVSYIAVATIFFNSGLSLKTEELTSALVHLKLHLFIQIF
TLAFFPATIWLFLQLLSITPINEWLLKGLQTVGCMPPPVSSAVILTKAVGGNEAAAIFNSAFGSFLGIVITPLLLLFL
GSSSSVPFTSIFSQLFMTVVVPLIIGQIVRRYIKDWLERKKPPFGAISSSVLLMIIYTTFCDTFSNPNIDLDKFSLVLI
LFIIFSIQLSFMLLTFIFSTRNNSGFTPADTVAIIFCSTHKSLTLGIPMLKIVFAGHEHLSLISVPLLIYHPAQILLGS
VLVPTIKSWMVSRQKGVKLTRPTV

3. 膜蛋白的跨膜区预测


地址(D) http://www.expasy.org/tools/#primary
输入中文，直接搜索
转到
YAHOO! Dnastar download 搜索 上网助手 清理 已拦截:617 杀毒 相册 购物 霍元甲
地址(D) http://www.cbs.dtu.dk/services/TMHMM-2.0/
EVENTS
NEWS
RESEARCH GROUPS
CBS PREDICTION SERVERS
CBS DATA SETS
PUBLICATIONS
BIOINFORMATICS EDUCATION PROGRAM
STAFF
CONTACT
INTERNAL
CBS BIOINFORMATICS TOOLS
OTHER BIOINFORMATICS LINKS
CBS >> CBS Prediction Servers >> TMHMM
TMHMM Server v. 2.0
Prediction of transmembrane helices in proteins
Update Nov. 29 2001: Minor change to the html output.
NOTE: You can submit many proteins at once in one fasta file. Please limit each submission to at most 4000 proteins. Please tick the 'One line per protein' option. Please leave time between each large submission.
Instructions
SUBMISSION
Submission of a local file in FASTA format (HTML 3.0 or higher)
浏览...
OR by pasting sequence(s) in FASTA format:
MRLLERMRKDWFMVGIVLAIAGAKLEPSIGVNGGPLKPEITVSYIAVATIFFNSGLSLKTEELTSALVH
LKLHLFIQIFTLAFFPATIWLFLQLLSITPINEWLLKGLQTVGCMPPPVSSAVILTKAVGGNEAAAIFN
SAFGSFLGIVITPLLLLLFLGSSSSVPFTSIFSQLFMTVVVPLIIGQIVRRYIKDWLERKKPPFGAISS
SVLLMIIYTTFCDTFSNPNIDLDKFSLVLILFIIFSIQLSFMLLTFIFSTRNNSGFTPADTVAIIFCST
HKSLTLGIPMLKIVFAGHEHLSLISVPLLIYHPAQILLGSVLVPTIKSWMVSRQKGVKLTRPTV

4. 蛋白质序列的二级结构预测



5. 蛋白质序列的三级结构预测

PEAt1蛋白的生物信息学分析

*pEAt1*基因序列（共624bp）

ATGGCCAACCCCCCGCATTGAAGAGCTCCCCGACGAGCCCGAGAAGAAGAACGTCCAG
ATCGACGACGATGAGTCCAGCGACGAGTCTGAGGGCGAGGAGGGCGAGGTCACCGT
ACCCCGCGGGCTCCTCCGTCGCTGTCCACTCCCGCAACGAGAAGAAGGCTCGCAAGGC
CATCGCCAAGCTCGGCCTGAAGCACATCGACGGCATCACACGCGTCACCCTCCGCCG
ACCCAAGAACATCCTCTTTGTCATCAACCAGCCCGACGTCTACAAGTCCCCTTCAAGCA
ACACCTGGATCATCTTCGGTGAGGCCAAGATCGAGGACCTCAACTCCCAGGCTCAGGC
TTCCGCCGCCCAGCAGCTTGCTCAGGCCGAGGCCGCATCCCACGACCACGCCGGCCA
CGACCACGGCGACGAGGCCAGCAAGGGCAAGGGCAAGGCTGTCGAGGACAAGAAGG
ACGAGGAGGAGGAGGATGACGATGAGGAGATTGACGACTCTGGCCTTGAGGCCAAGG
ACATCGAGCTCGTCATGCAGCAGGCCAGCGTTTCGCGGAAGAAGGCCGTCAAGGCCC
TCAAGGAGAACGATAACGATATAGTCAACTCCATCATGGCGCTGAGCATATAG

PEAt1蛋白序列（207aa）

MANPRIEELPDEPEKKNVQIEEDESSDESEGEEGEVSVPAGSSVAVHSRNEKKARKAIAKL
GLKHIDGITRVTLRRPKNILFVINQPDVYKSPSSNTWIIFGEAKIEDLNSQAQASAAQQLAQA
EAASHDHAGHDHGDEASKGKGKAVEDKKDEEEEDDDEEIDDSGLEAKDIELVMQQASVSR
KKAVKALKENDNDIVNSIMALSI

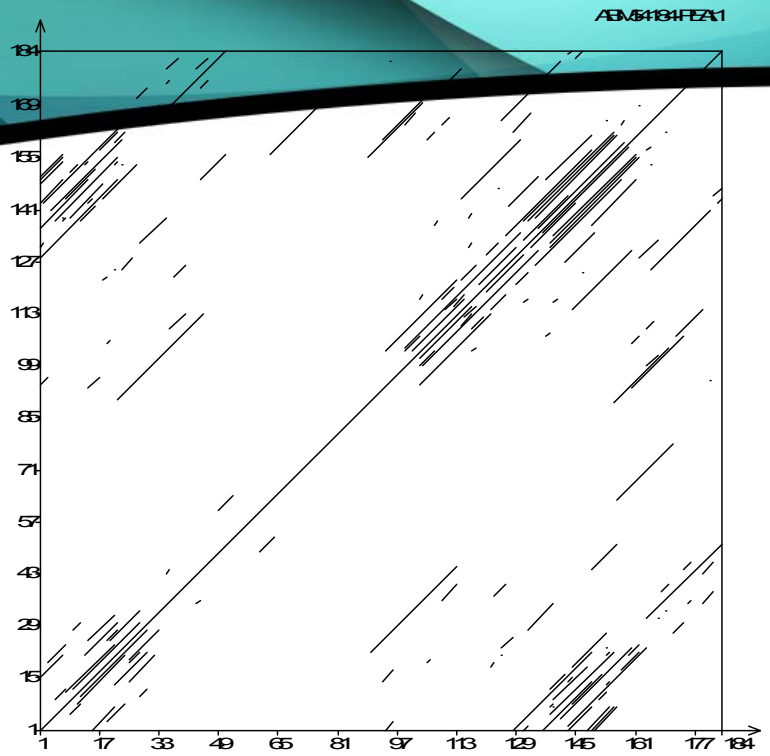
PEAt1序列的初步分析

- Molecular weight = 22632.80
- Residues = 207
- Average Residue Weight = 109.337
- Charge = -16.0
- Isoelectric Point = 4.4305
- A280 Molar Extinction Coefficient = 6970
- A280 Extinction Coefficient 1mg/ml = 0.31

Residue	Number	Mole%	DayhoffStat
A = Ala	23	11.111	1.292
B = Asx	0	0.000	0.000
C = Cys	0	0.000	0.000
D = Asp	19	9.179	1.669
E = Glu	27	13.043	2.174
F = Phe	2	0.966	0.268
G = Gly	11	5.314	0.633
H = His	6	2.899	1.449
I = Ile	15	7.246	1.610
J = ---	0	0.000	0.000
K = Lys	20	9.662	1.464
L = Leu	11	5.314	0.718
M = Met	3	1.449	0.853
N = Asn	10	4.831	1.123
O = ---	0	0.000	0.000
P = Pro	7	3.382	0.650
Q = Gln	9	4.348	1.115
R = Arg	7	3.382	0.690
S = Ser	19	9.179	1.311
T = Thr	3	1.449	0.238
U = ---	0	0.000	0.000
V = Val	13	6.280	0.952
W = Trp	1	0.483	0.372
X = Xaa	0	0.000	0.000
Y = Tyr	1	0.483	0.142

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	56	27.053
Small	(A+B+C+D+G+N+P+S+T+V)	105	50.725
Aliphatic	(A+I+L+V)	62	29.952
Aromatic	(F+H+W+Y)	10	4.831
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	87	42.029
Polar	(D+E+H+K+N+Q+R+S+T+Z)	120	57.971
Charged	(B+D+E+H+K+R+Z)	79	38.164
Basic	(H+K+R)	33	15.942
Acidic	(B+D+E+Z)	46	22.222

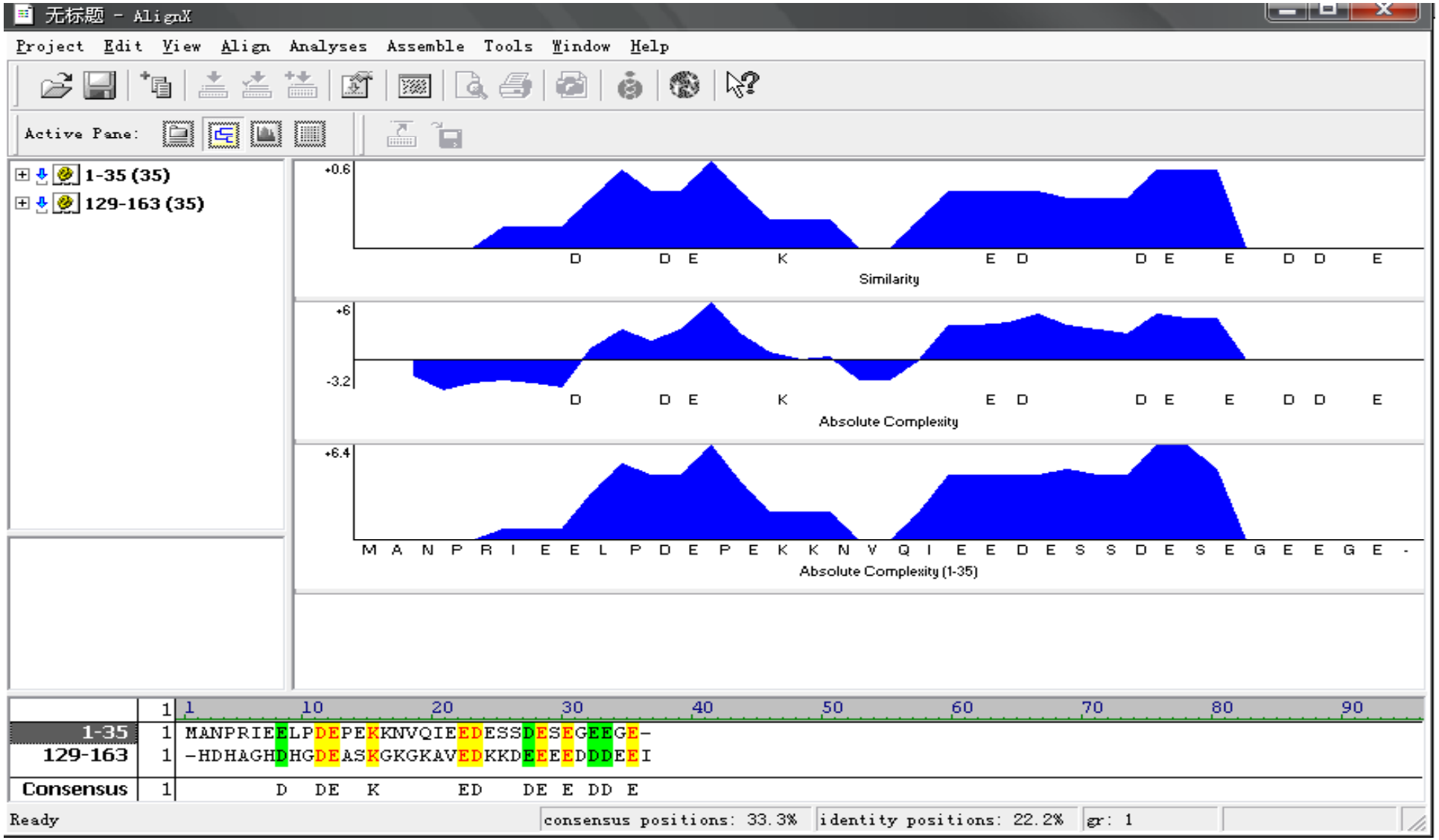
Self alignment of PEAt1



所用软件: Vector NTI Advance 10软件包的Dot Matrix

参数: stringency 20; window 24

从上图可看出PEAt1氨基酸序列的1-35残基同129-163残基具有显著相似性，极性特别是酸性氨基酸出现的频率很高。推测这两段序列可能以某种方式出现蛋白表面，构成富含负电荷的亲水侧。

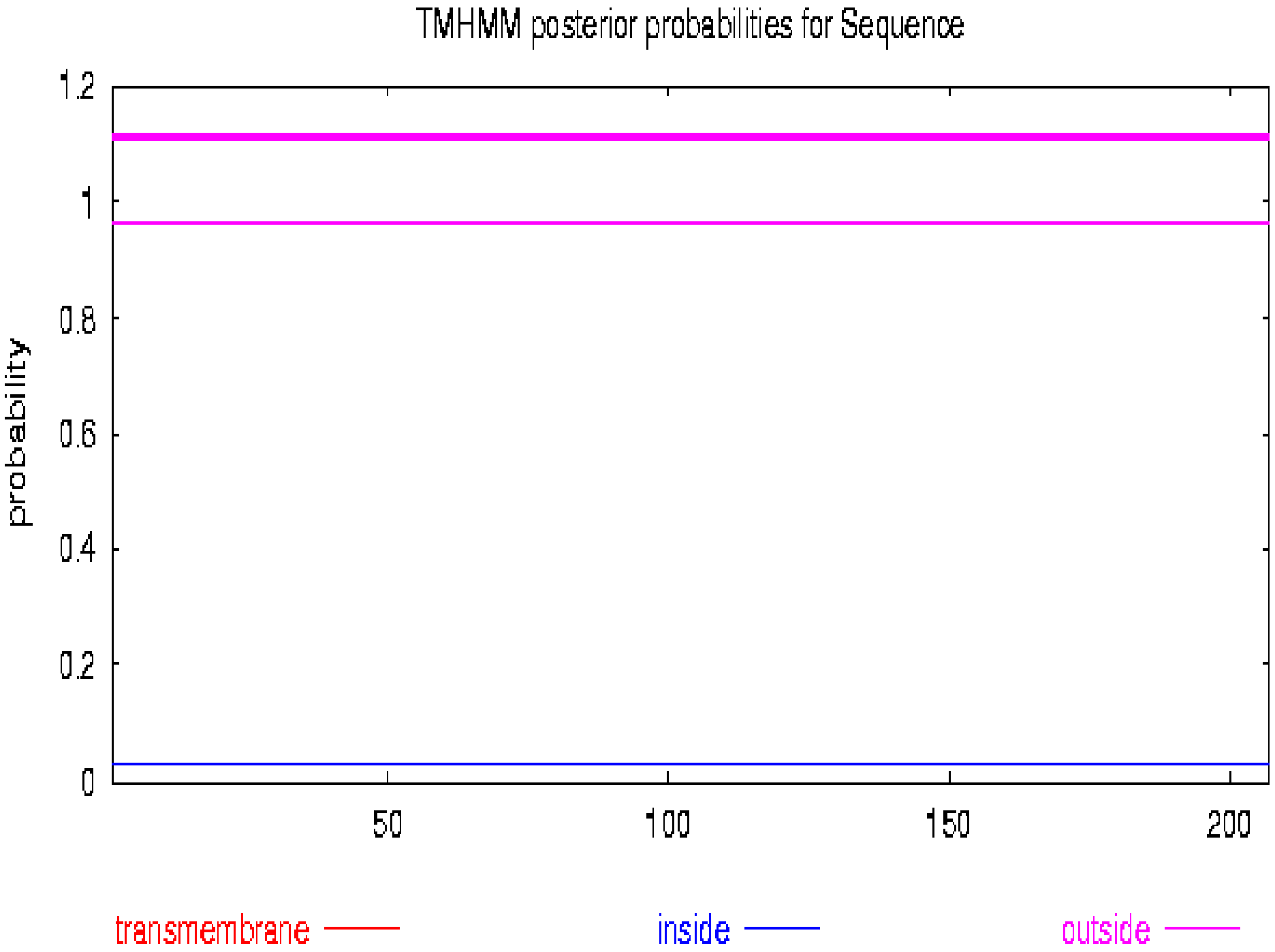


进一步分析：这两段序列的一致性为22.2%，相似性为33.3%

跨膜区预测

TMHMM2.0

- PEAt1
- # Sequence Length: 207
- # Sequence Number of predicted TMHs: 0
- # Sequence Exp number of AAs in TMHs: 0.00038
- # Sequence Exp number, first 60 AAs: 4e-05
- # Sequence Total prob of N-in: 0.03326
- Sequence TMHMM2.0 outside 1-207


TMHMM posterior probabilities for Sequence
probability
1.2
1
0.8
0.6
0.4
0.2
0
50
100
150
200
transmembrane
inside
outside

does it contain transmembrane segments?

- Report_format: seqtable

Report_file: peat1.tmap

#==

Sequence: Peat1 from: 1 to: 207

HitCount: 0

#==

Start End TransMem Sequence

信号肽预测

SignalP 3.0 Server - prediction results

SignalP-NN (neural networks) **result:**

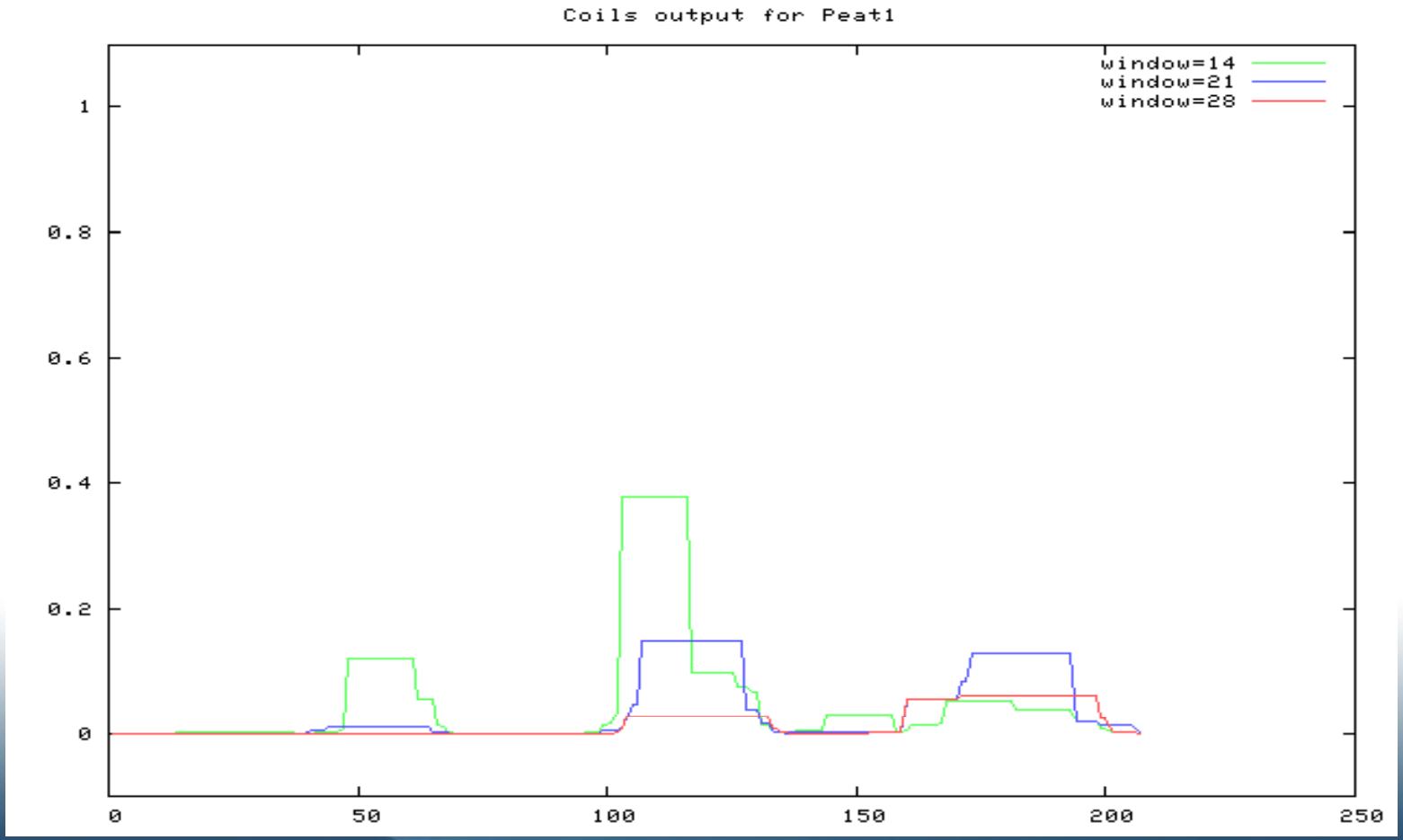
>seq1 length = 70

Measure	Position	Value	Cutoff	signal peptide?
max. C	32	0.060	0.32	NO
max. Y	32	0.018	0.33	NO
max. S	1	0.016	0.87	NO
mean S	1-31	0.009	0.48	NO
D	1-31	0.014	0.43	NO

Does it contain coiled-coils?

NCOILS version 1.0

using MTIDK matrix, weights: a,d=2.5 and b,c,e,f,g=1.0



➔Secondary Structure Prediction

- PROFsec summary

overall your protein can be classified as:
>>> mixed <<<
given the following classes:
\all-alpha\: %H > 45% AND %E < 5%
\all-beta\: %H < 5% AND %E > 45%
\alpha-beta\: %H > 30% AND %E > 20%
\mixed\: all others
Predicted secondary structure composition:

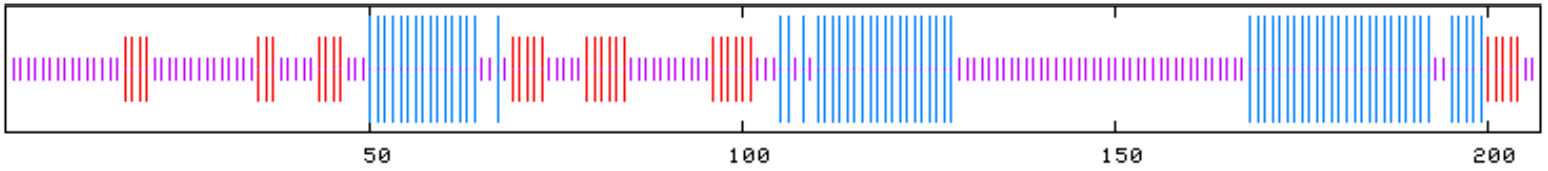
sec str type	H	E	L
% in protein	29.5	10.6	59.9

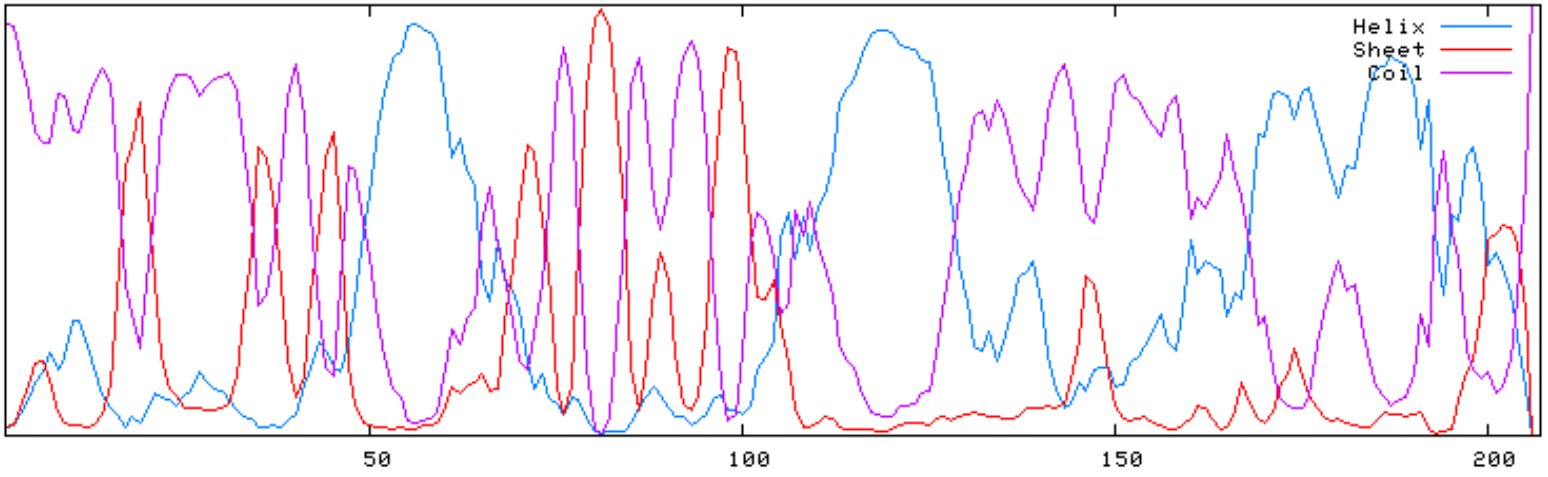
Predicted solvent accessibility composition (core/surface ratio):
Classes used:
e: residues exposed with more than 16% of their surface
b: all other residues.

acc type	b	e
% in protein	22.2	77.8

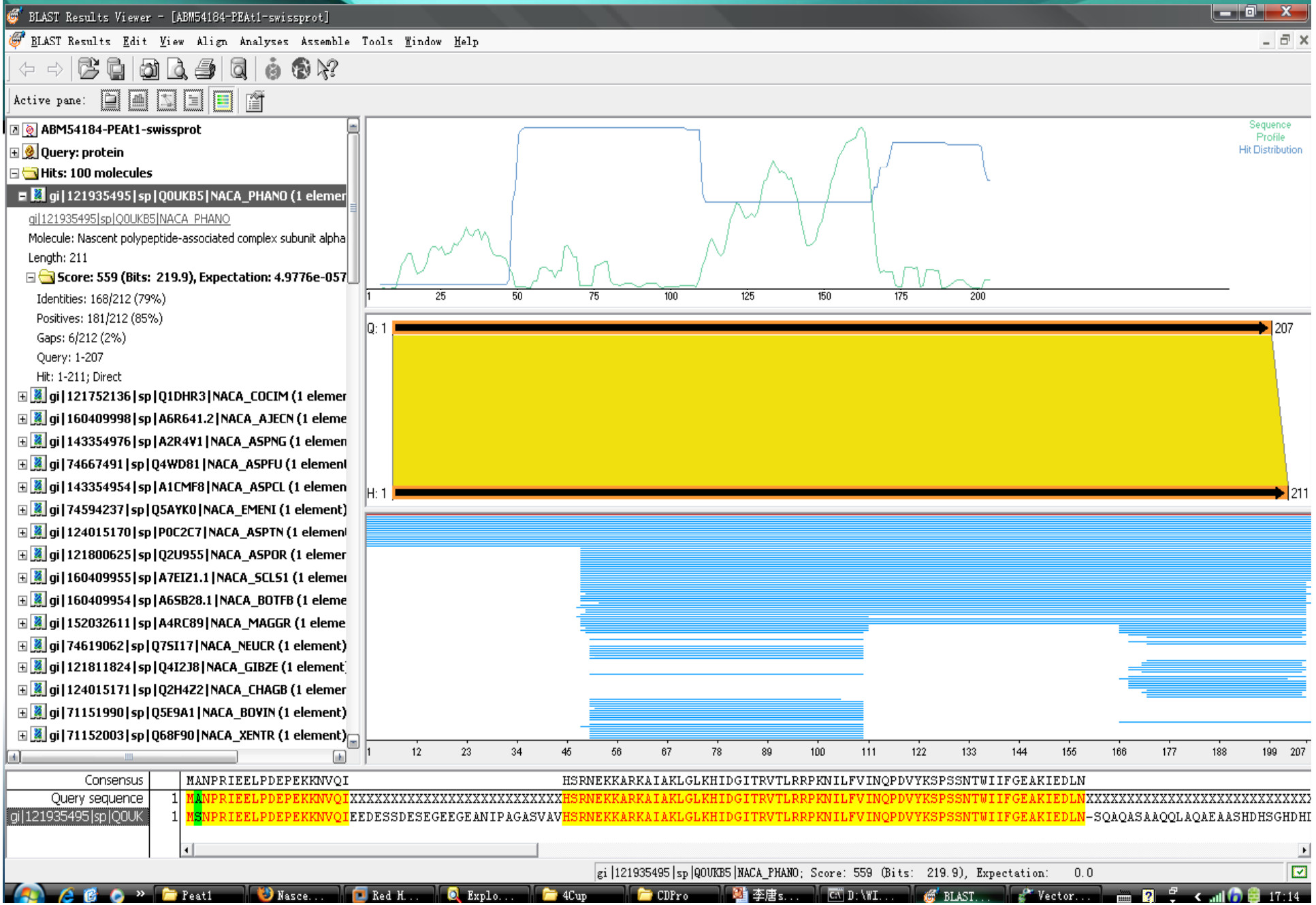
HNN - Hierarchical Neural Network method

Alpha helix	(Hh)	:	68	is	32.85%
3_{10} helix	(Gg)	:	0	is	0.00%
Pi helix	(Ii)	:	0	is	0.00%
Beta bridge	(Bb)	:	0	is	0.00%
Extended strand	(Ee)	:	33	is	15.94%
Beta turn	(Tt)	:	0	is	0.00%
Bend region	(Ss)	:	0	is	0.00%
Random coil	(Cc)	:	106	is	51.21%
Ambigous states	(?)	:	0	is	0.00%
Other states		:	0	is	0.00%





➔Sequence Database searching



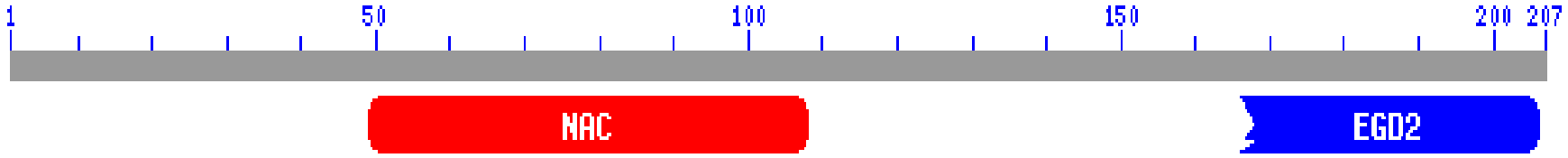
Peat1 Sequence Database searching

- 基因序列的同源比对发现与NAC蛋白高度同源：

  Identities = 426/516 (82%)

- NAC编码基因的102-614位序列与PEAt1的108-623位序列相匹配，有82%的一致性

- 蛋白序列比对发现与NAC蛋白家族结构域NAC和EGD2（UBA）有高度同源。



- **Q0UKB5** (NACA_PHANO)

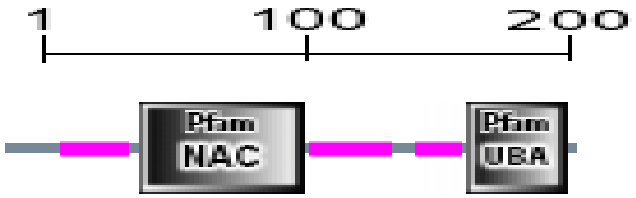
Nascent polypeptide-associated complex subunit alpha

msnprieelpdepekknvqieedessdesegeegeanipagasvavhsrnekkarkaia
klglkhidgitrvtlrrpknilfvinqpdvykspssntwiifgeakiedlnsqaqasaaqqlaqae
aashdhsghdhdhdhkgkgkavedkkddeedeeeeeeddeevddsgleakdielvm
qqasvsrkkavkalkendndivnsimalsi

Chain	1 - 211	211	Nascent polypeptide-associated complex subunit alpha	

ons

Domain	48 - 113	66	NAC-A/B	
Domain	172 - 211	40	UBA	



NAC Domain

Component of the nascent polypeptide-associated complex (NAC), a dynamic component of the ribosomal exit tunnel. Functions:

1. protect the emerging polypeptides from interaction with other cytoplasmic proteins to ensure appropriate nascent protein targeting.
2. promotes mitochondrial protein import by enhancing productive ribosome interactions with the outer mitochondrial membrane and blocks the inappropriate interaction of ribosomes translating non-secretory nascent polypeptides with translocation sites in the membrane of the endoplasmic reticulum.
3. Involve in transcription regulation.

NAC Domain

- 保护新生多肽不与胞质内其它蛋白不正确聚集;
- 确保新生肽链的靶向转运;
- 促进线粒体蛋白进入线粒体;
- 阻止非分泌新生多肽发生在内质网膜上的跨膜转移;
- 参与转录调控

UBA Domain

The ubiquitin-associated (UBA) domain is an approximately 40 amino acid motif that was first recognized in proteins associated with ubiquitination but is also found in proteins involved in DNA nucleotide excision-repair and other proteins. UBA domains have been shown to bind mono-, di-, tri-, and tetra-ubiquitin *in vitro* but appear to bind to polyubiquitin with a higher affinity and it is thought that polyubiquitinated proteins represent the true *in vivo* binding substrates. As well, some UBA domains appear to homo and heterodimerize and to bind other substrates. **Functionally**, the UBA domain has been proposed to limit ubiquitin chain elongation and to target ubiquitinated proteins to the 26S proteasome for degradation.

UBA Domain

- 最先因为其同泛素结合而被认识，参与DNA切除修复的蛋白中也发现其存在。可同多种聚合状态的泛素结合－多聚泛素化，一些蛋白的UBA常形成同或异二聚体或同其它物质结合;

- 功能上主要是通过泛素化降解蛋白质

总结与讨论

对于Peat1抗逆，抗病，促进生长的解释

- 通过UBA结构域参与泛素降解途径，清除体内胁迫产生的多余蛋白质。
- 通过NAC结构域协助新生多肽正确折叠。

➔Comparative or homology modeling

- Comparative modeling steps:

Step 1: Searching for structures related to the target sequence

Step 2: Selecting templates

Step 3: Aligning the target sequence with one or more structures

Step 4: Model Building

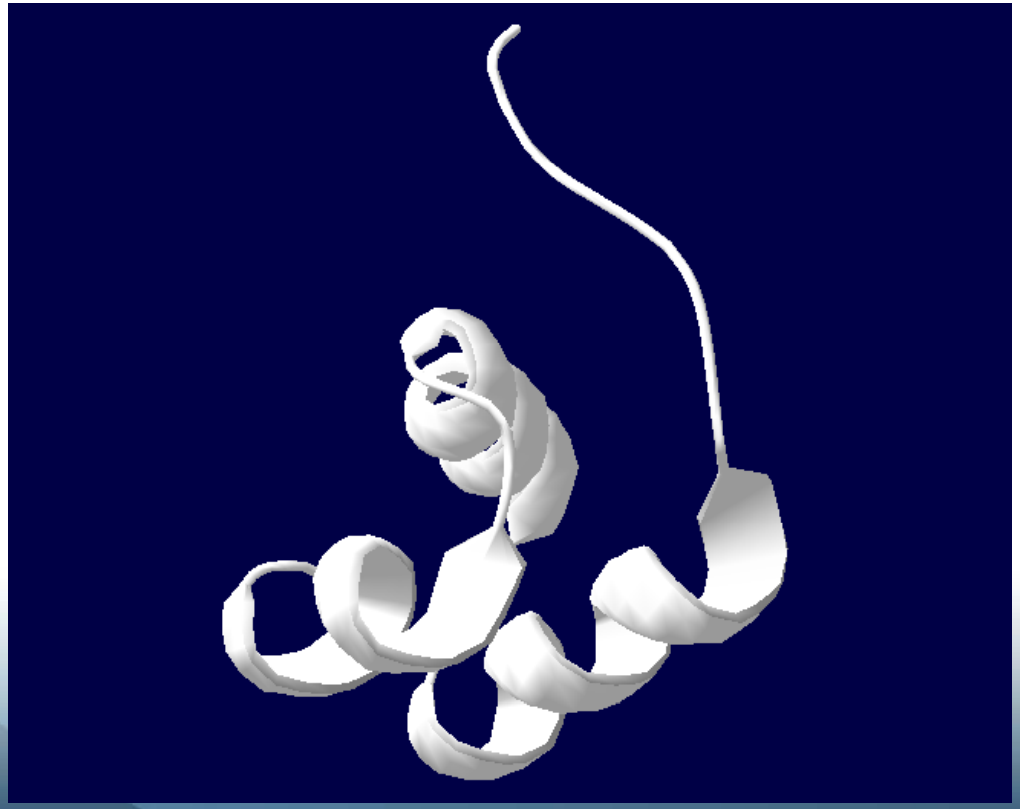
Step 5: Evaluating a model

同源建模

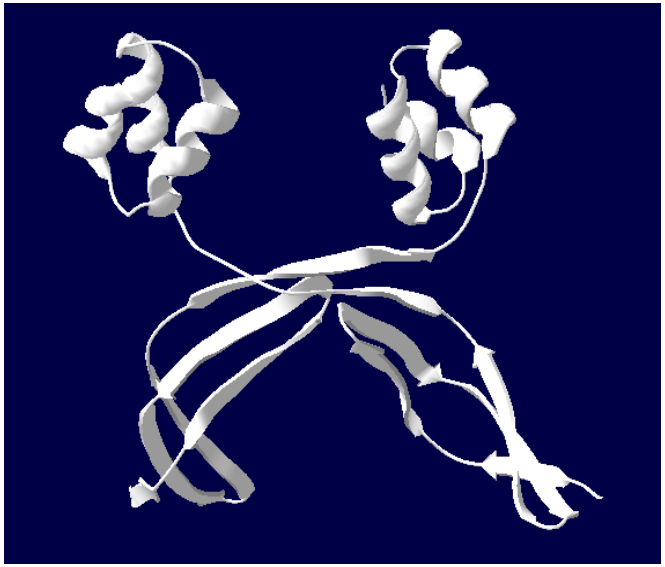
- 将Peat1序列到蛋白三维结构数据库（PDB）进行比对，寻找合适的模板
- 将得到的模板同Peat1序列进行比对
- 根据模板和Peat1序列间的匹配关系，按照模板序列各氨基酸残基间的定位及约束关系构建Peat1序列的相应三维模型，并对模型进行优化
- 模型的评价：计算模型的能量状态，各残基的立体排布和走向的合理性

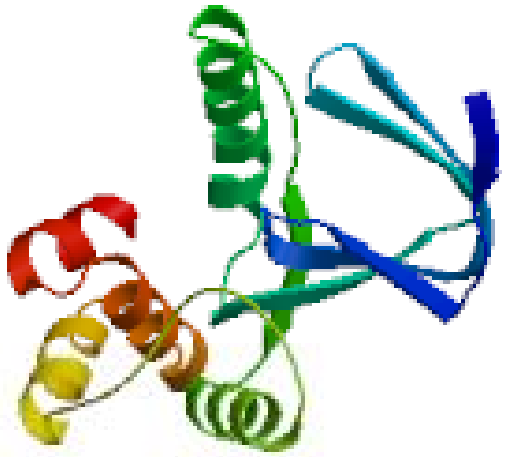
>PEAt1 | 163-206
IDDSGLEAKDIELVMQQASVSRKKAVKALKENDNDIVNSIMALS

UTHOR： ESyPred3D server
PARENT： 1TR8_A
>1TR8-Am | 76-116 | UBA Domain
EXEIPEDDIELVXNQTGASREDATRALQETGGDLAEAIXRL

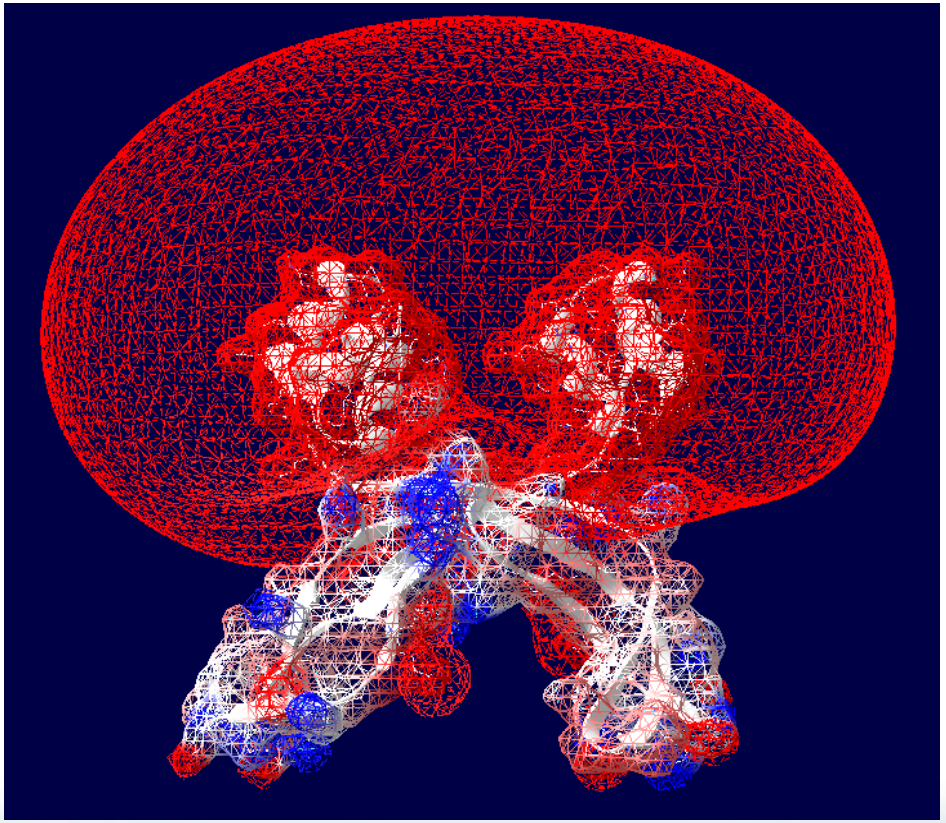


1TR8同二聚体



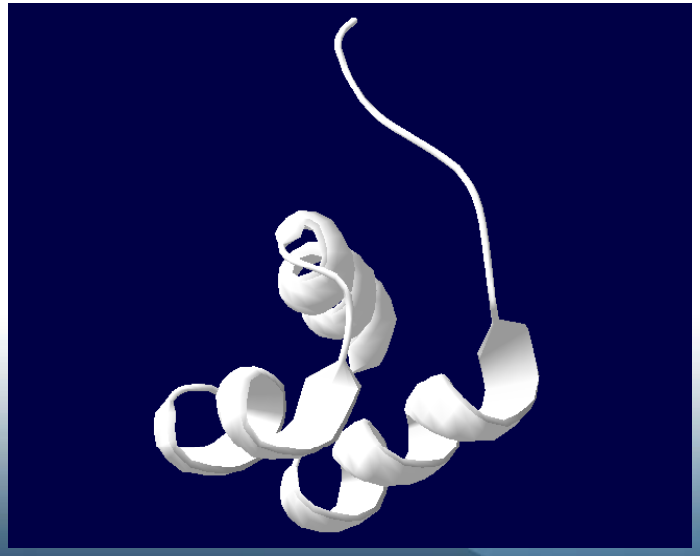


1TR8的表面结构和电荷分布



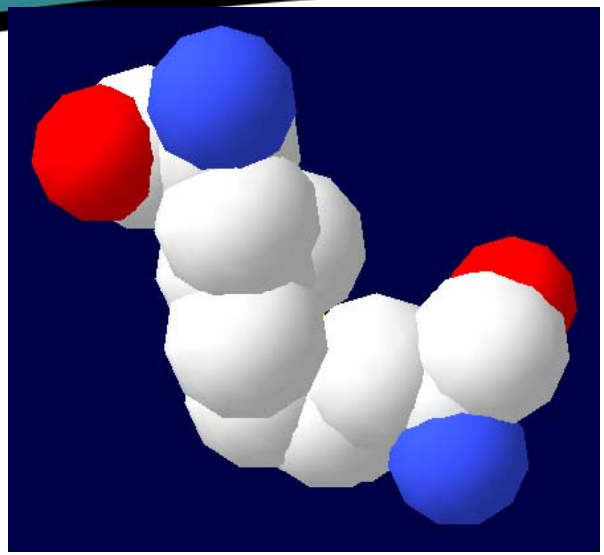
提交Peat1序列到SWISS MODEL 服务器进行同源建模，有一小部分PEAT1的结构可以得到。

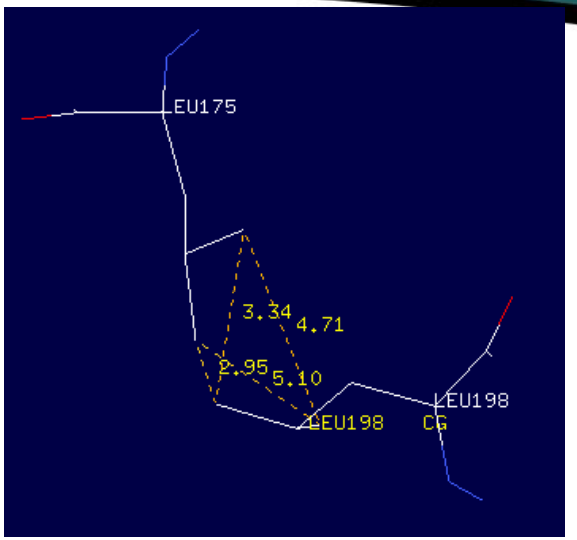
```
1tr8A     1 DMKDLRGVEEVVIKL.KRKEIIIKNPKVNVMEFMGQKTYQVTGKARERSL.. 49
              K   G   V     K     I  P V         T    G A    L
Target   62 GLKHIDGITRVTLRRpKNILFVINQPDV--YKSPSSNTWIIFGEAKIEDLns 111

1tr8A    50 .....................................................E 50

Target  112 qaqasaaqqlaqaeaashdhaghdhgdeaskgkgkavedkkdeeeedddeeI 163

1tr8A    51 AEMEIPEDDIELVMNQTGASREDATRALQETGGDLAEAIMRL 92
                    DIELVM Q   SR  A  AL E   D    IM L
Target  164 DDSGLEAKDIELVMQQASVSRKKAVKALKENDNDIVNSIMAL 205
```



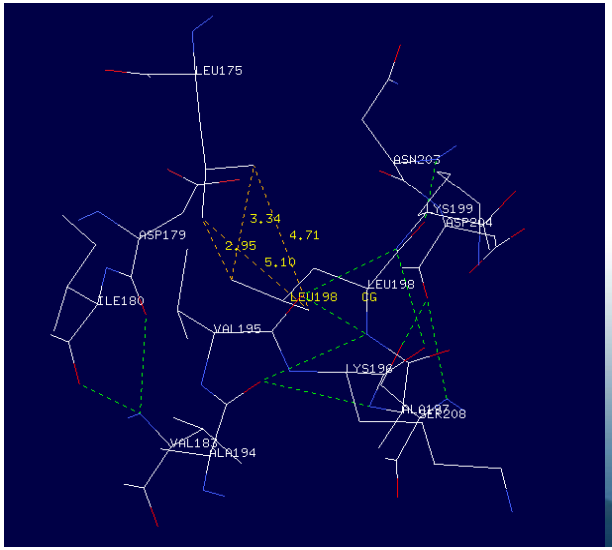
以1tr8A为模板，PEAt1的165-205位氨基酸的预测结构。

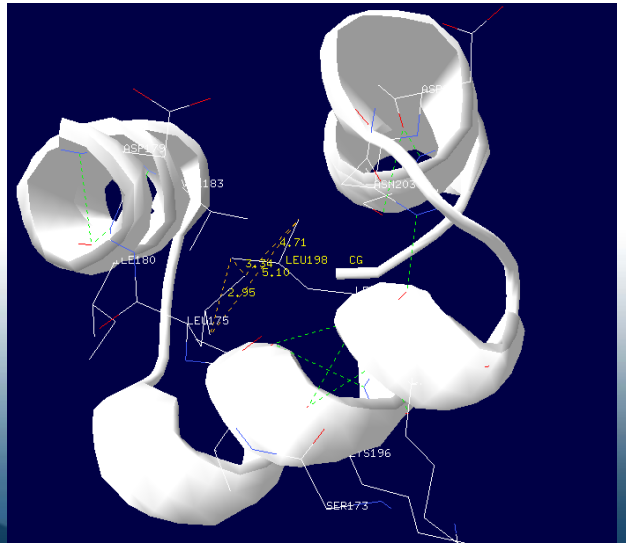
PEAt1-UBA Domain的疏水核心：螺旋1N端的Leu175同Leu198的疏水侧链结合，Leu175侧链包埋疏水核心，使螺旋1终止。

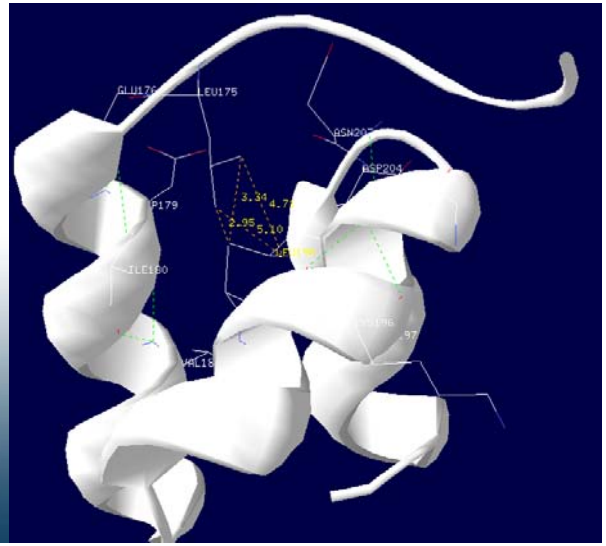


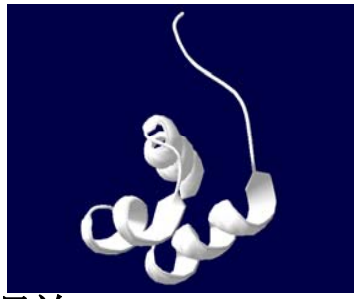


左：侧视图　　中：顶视图　　右：侧视图(模型)



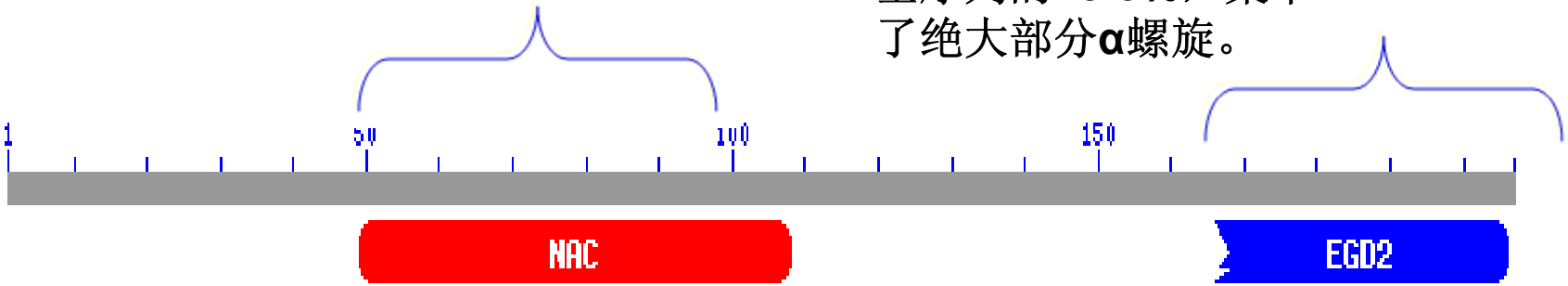


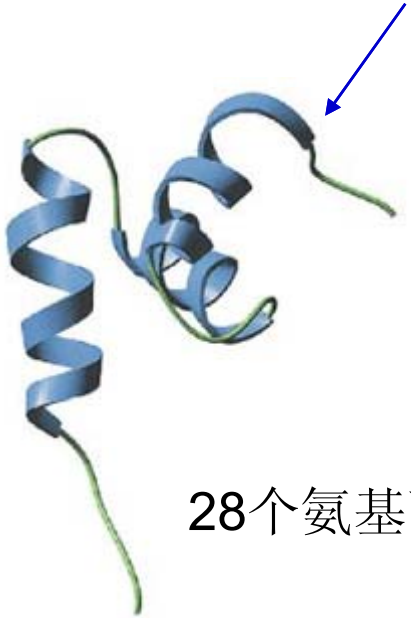




a-Helix b-Sheet
16.1% 30.7%

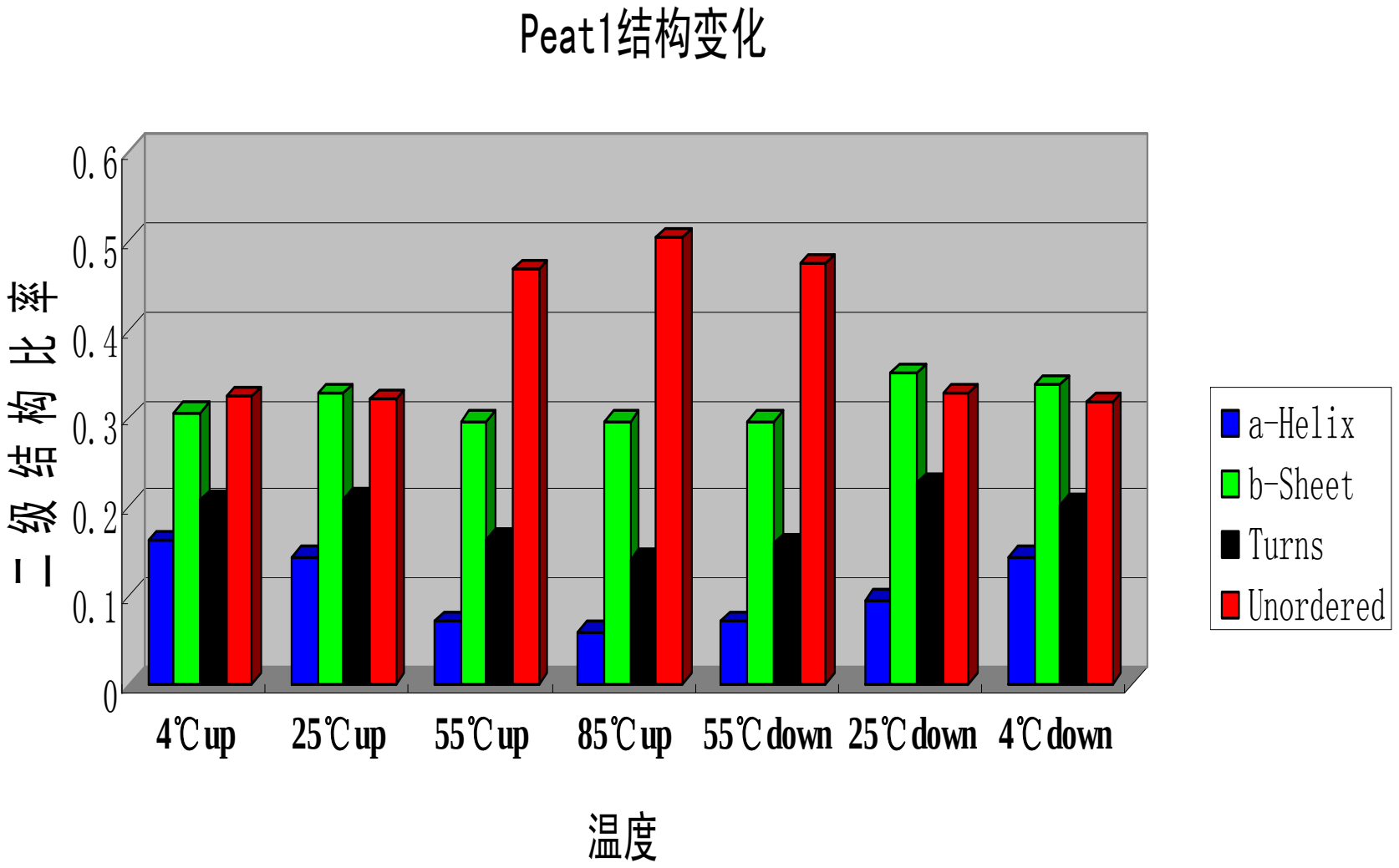
28个氨基酸的α螺旋，全序列的13.5%，集中了绝大部分α螺旋。





28个氨基酸的α螺旋

CD（圆二色谱）分析


Peat1结构变化
二级结构比率
0.6
0.5
0.4
0.3
0.2
0.1
0
4℃up
25℃up
55℃up
85℃up
55℃down
25℃down
4℃down
温度
a-Helix
b-Sheet
Turns
Unordered

- 首先，得到了Peat1蛋白在正常温度下的二级结构构成。
- Peat1蛋白随着温度升高，β折叠变化不大，可以认为是维持结构稳定的骨架结构。α螺旋结构减少，相应的转角结构也随之减少，这些结构都变为无规则卷曲。
- 随着温度的恢复，α螺旋能够逐渐恢复为原来水平。
- Peat1蛋白的耐热机制应为结构可恢复型。

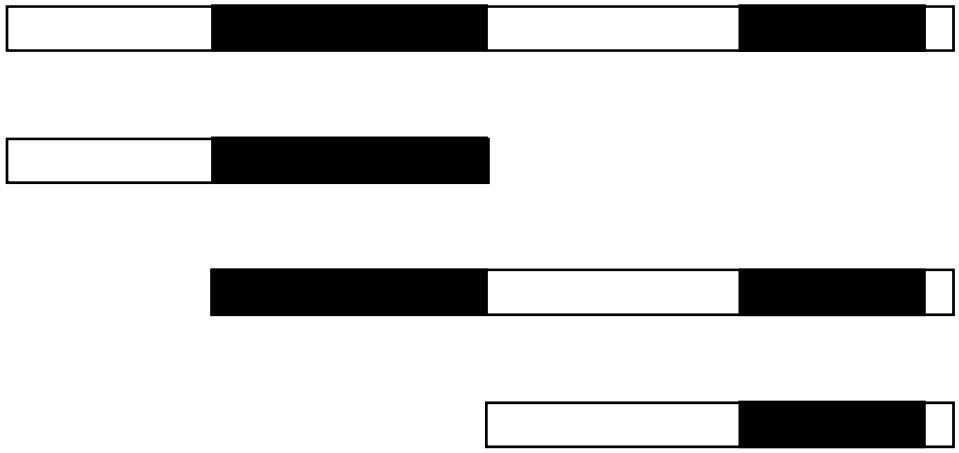
缺失突变

Peat1

Peat1-△CD99

Peat1-△ND49

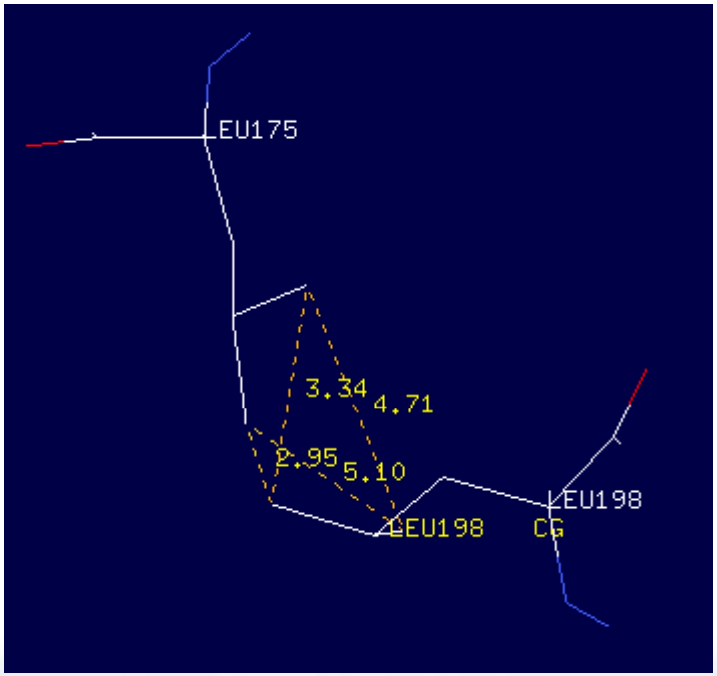
Peat1-△ND108



这几个突变体都具有耐热的特性

总结与讨论

- β折叠是维持结构稳定的因素之一



- 疏水作用维持的α螺旋是耐热的另一个因素

175位的LEU和198位的LEU侧链距离只有2.95埃，疏水作用强，推测是维持α螺旋核心稳定的因素之一，在高温时，α螺旋也仍然存在6%。

总结与讨论

对于Peat1抗逆，抗病，促进生长的解释

- 通过UBA结构域参与泛素降解途径，清除体内胁迫产生的多余蛋白质。
- 通过NAC结构域协助新生多肽正确折叠。

总结与讨论

- Peat1蛋白的生物信息学分析结果主要为新型激活蛋白防虫、抗病、促生机制提供了理论基础。为下一步深入解析Peat1蛋白的高级结构及其作用机理提供了可行的切入点。

Acknowledgement

- Mr.Luo
- Group members
- Li T

References

- http://abc.cbi.pku.edu.cn/
- http://cn.expasy.org/
- Jin xiong.Essential Bioinformatics. 2006，Cambridge University Press.
- 孙啸，陆祖宏等.生物信息学基础.2006，清华大学出版社

Half day on the web,

saves you

half month in the lab.

——Alan Bleasby

thanks !