

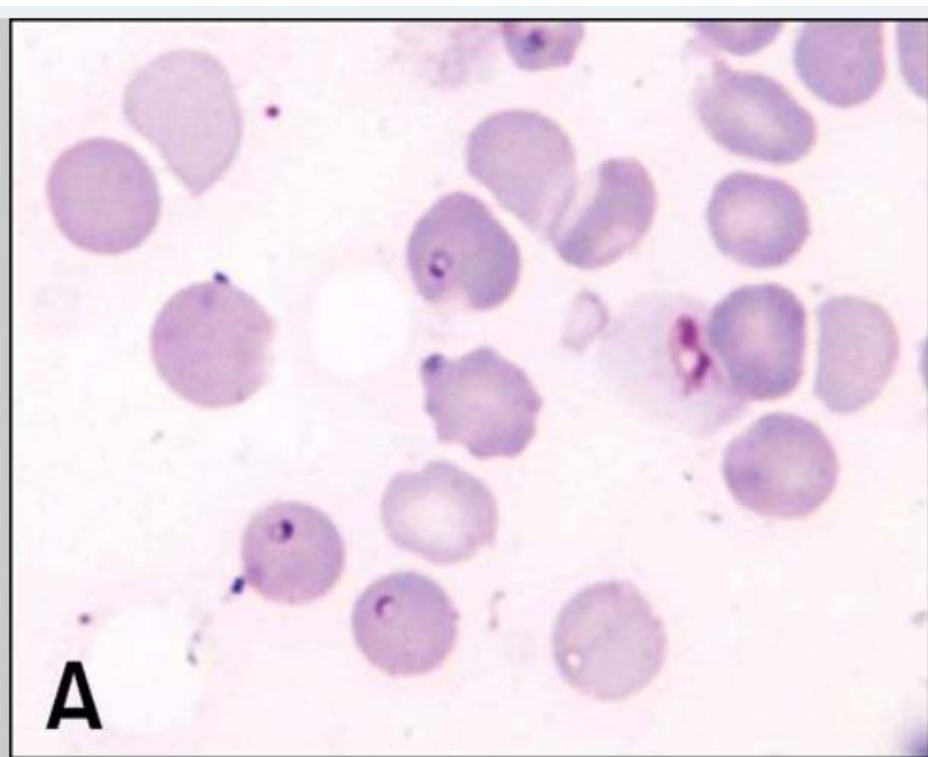
# 环形泰勒虫TASP蛋白抗原性分析



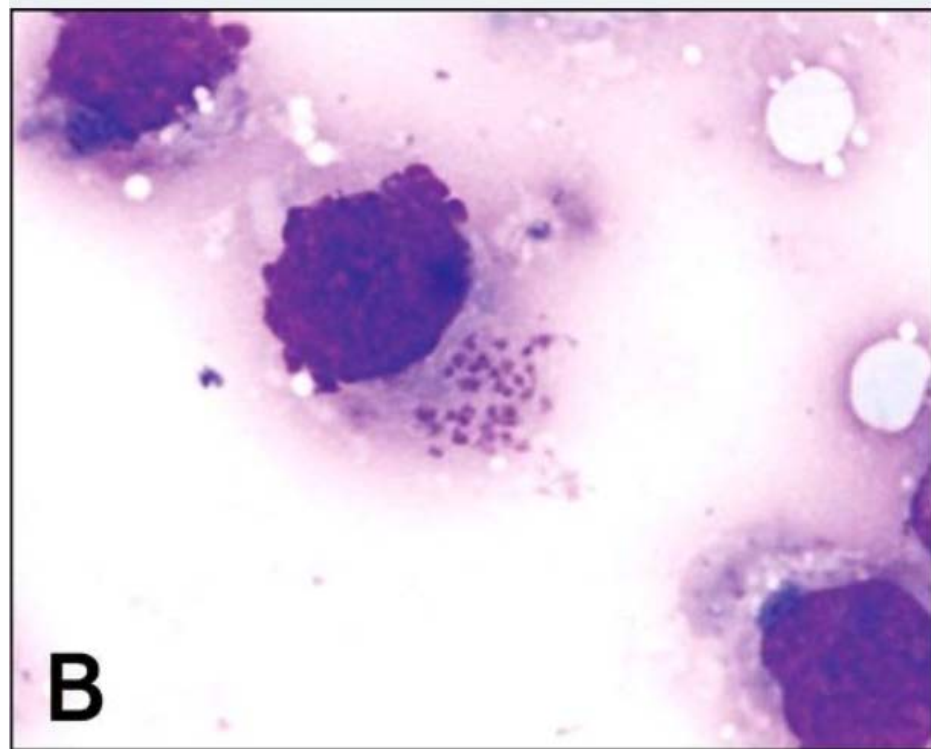
报告人：卢怡竹

组员：张伟、孙铭苑

报告时间：2013年6月21日



A 红细胞中的泰勒虫



B 有核细胞中裂殖体

## 研究背景

泰勒虫寄生于红细胞以及淋巴细胞内，引起以高热、贫血、出血、消瘦和体表淋巴结肿大为主要临床症状的血液原虫病。随着养殖业的不断发展，泰勒虫病的发生也呈上升趋势，严重影响国家的畜牧业发展。对环形泰勒虫TaSP 基因抗原性研究，可为研制泰勒虫病疫苗和诊断试剂盒提供科学依据。

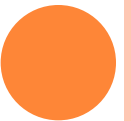


- 蛋白质特性
  - 疏水性
  - 亲水性
  - 信号肽
  - 亚细胞定位
- 蛋白质结构
  - 一级结构
  - 二级结构



NCBI登陆号：XM\_947650.1，基因全长945bp。

Uniprot登录号：Q8WPH8，编码314个氨基酸。



# 蛋白质一级结构分析

分析内容：

氨基酸组成

等电点(pI)

分子量 (molecular weight)

疏水性 (hydrophobicity)



PEPSTATS of Q8WPH8\_THEAN from 1 to 314

Molecular weight = 35449.97      Residues = 314  
 Average Residue Weight = 112.898      Charge = -17.5  
 Isoelectric Point = 4.4314  
 A280 Molar Extinction Coefficient = 29870  
 A280 Extinction Coefficient 1mg/ml = 0.84  
 Probability of expression in inclusion bodies = 0.764

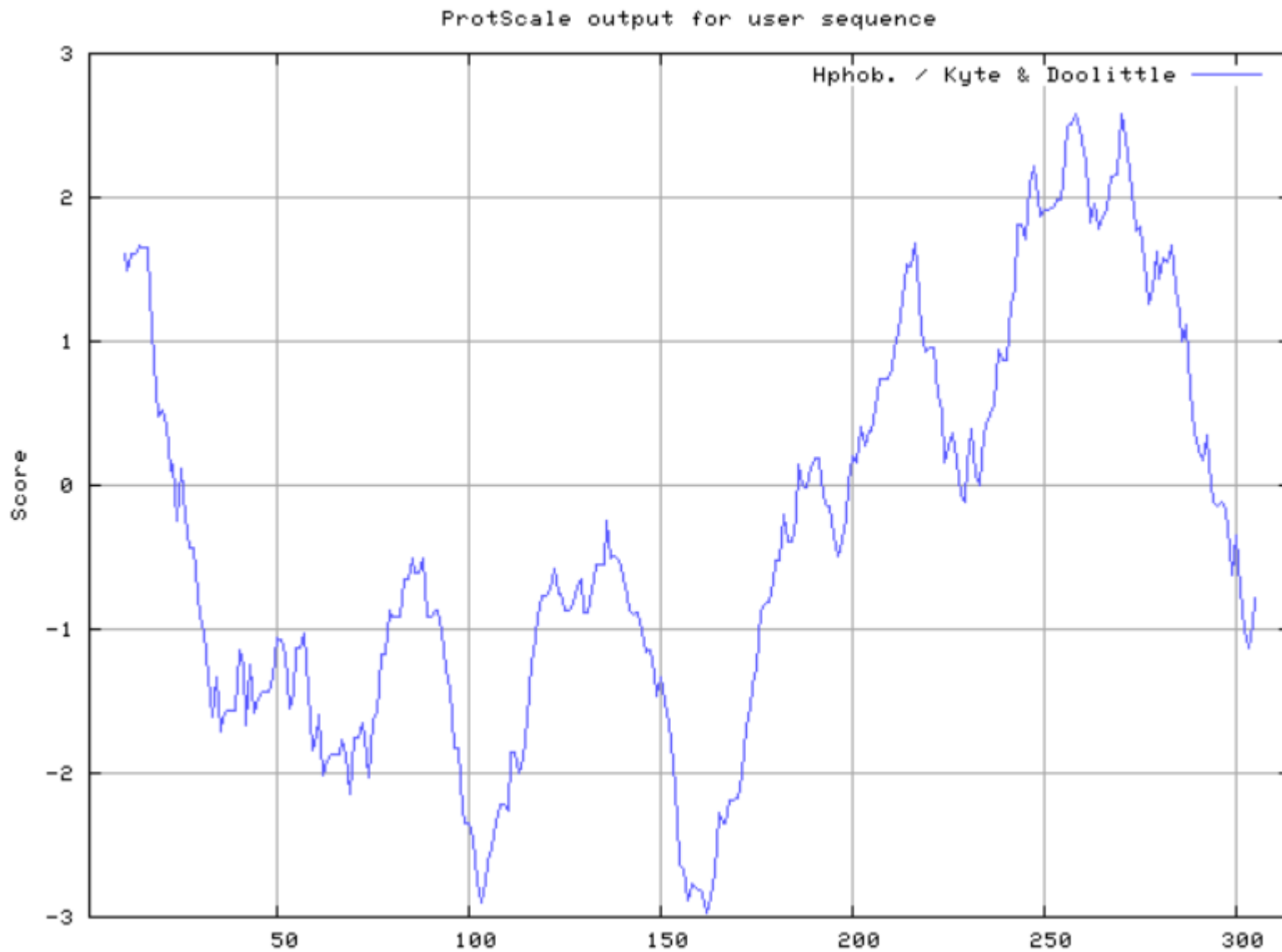
Residue	Number	Mole%	DayhoffStat
A = Ala	20	6.369	0.741
B = Asx	0	0.000	0.000
C = Cys	11	3.503	1.208
D = Asp	22	7.006	1.274
E = Glu	25	7.962	1.327
F = Phe	23	7.325	2.035
G = Gly	10	3.185	0.379
H = His	9	2.866	1.433
I = Ile	11	3.503	0.778
J = ---	0	0.000	0.000
K = Lys	17	5.414	0.820
L = Leu	30	9.554	1.291
M = Met	3	0.955	0.562
N = Asn	5	1.592	0.370
O = ---	0	0.000	0.000
P = Pro	30	9.554	1.837
Q = Gln	15	4.777	1.225
R = Arg	8	2.548	0.520
S = Ser	26	8.280	1.183
T = Thr	18	5.732	0.940
U = ---	0	0.000	0.000
V = Val	18	5.732	0.869
W = Trp	3	0.955	0.735
X = Xaa	0	0.000	0.000
Y = Tyr	10	3.185	0.937
Z = Glx	0	0.000	0.000

Property	Residues	Number	Mole%	
Tiny	(A+C+G+S+T)	85		27.070
Small	(A+B+C+D+G+N+P+S+T+V)	160		50.955
Aliphatic	(A+I+L+V)	79	25.159	
Aromatic	(F+H+W+Y)	45	14.331	
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	169	53.822	
Polar	(D+E+H+K+N+Q+R+S+T+Z)	145	46.178	
Charged	(B+D+E+H+K+R+Z)	81		25.796
Basic	(H+K+R)	34		10.828
Acidic	(B+D+E+Z)	47		14.968

## Pepstats统计结果

分子量为35449.97 ku，  
等电点为4.4314。





ProtScale预测 亲水性参数





# 蛋白质二级结构

分析内容：

$\alpha$ 螺旋

$\beta$ 折叠

$\beta$  转角

无规则卷曲



```

      . 10 . 20 . 30 . 40 . 50
MKFFYLFLVLPILLKFCECGPFLPLDRQLNPIDFDPNDQHPDPDQLID
helix HHHH HHHHHHH
sheet EE E E EEEE
turns TTTTTTTT TTTT TTT TTTTT TT
coil C CC C CCCC
      . 60 . 70 . 80 . 90 . 100
QIEPSEQPAQQEPIEPQQPTQPSTPEPEELQPETVTVEVPEPVTSEEPKES
helix HHHHHH HHHHH HH
sheet E EEEEE
turns T T TT TT T T
coil CCC CCC CCC C CCCCCCCC CCCC
      . 110 . 120 . 130 . 140 . 150
DQTEEQKHEEPEASPAPEPVDEPAVHATESTPTKASSSGDGAAVCHGKHH
helix HHHHHHHHHH HHHHHHHHHH HHHHH
sheet
turns T T TT TT TT T
coil CCCCCC CCC CCCCC C C
      . 160 . 170 . 180 . 190 . 200
DYDSDGKESKSDHKRPKDKKPFVPKTSQCCGSYFTNSYKITVAFDWWLC
helix HHHHHHHHHH
sheet E E EEEEE
turns TTTTTT TTTT T TTTTTTTTTTTTTTTT TTTTT
coil C
      . 210 . 220 . 230 . 240 . 250
DKPWQYALTLALFGFSLSPCLKAYREVLRAKAIRSFIFDCFLTHLFLF
helix HH HHH HHHHHHHHHHHH HHHHHHHH
sheet EEEEE E EEEE
turns T TT TTT TT TTT
coil C CC

```

```

      . 260 . 270 . 280 . 290 . 300
LIAFCAYALDFLLMLVMTFNVGVFFAVITGYTVGYLVSSLAYSTLRSH
helix HHHHHHHHHHHHHHHHHH
sheet EEE EEEEEEE EEEEEEEEEEE EEE
turns TT TT T T
coil CC CC CCC
      . 310
ARSSSFSRINEDCC
helix
sheet
turns TTT TTTTTTTTT
coil CC
#-----
#
# Residue totals: H:104 E: 54 T: 97 C: 59
# percent: H: 34.9 E: 18.1 T: 32.6 C: 19.8
#-----
#-----
# Total_sequences: 1
# Total_hitcount: 77
#-----

```

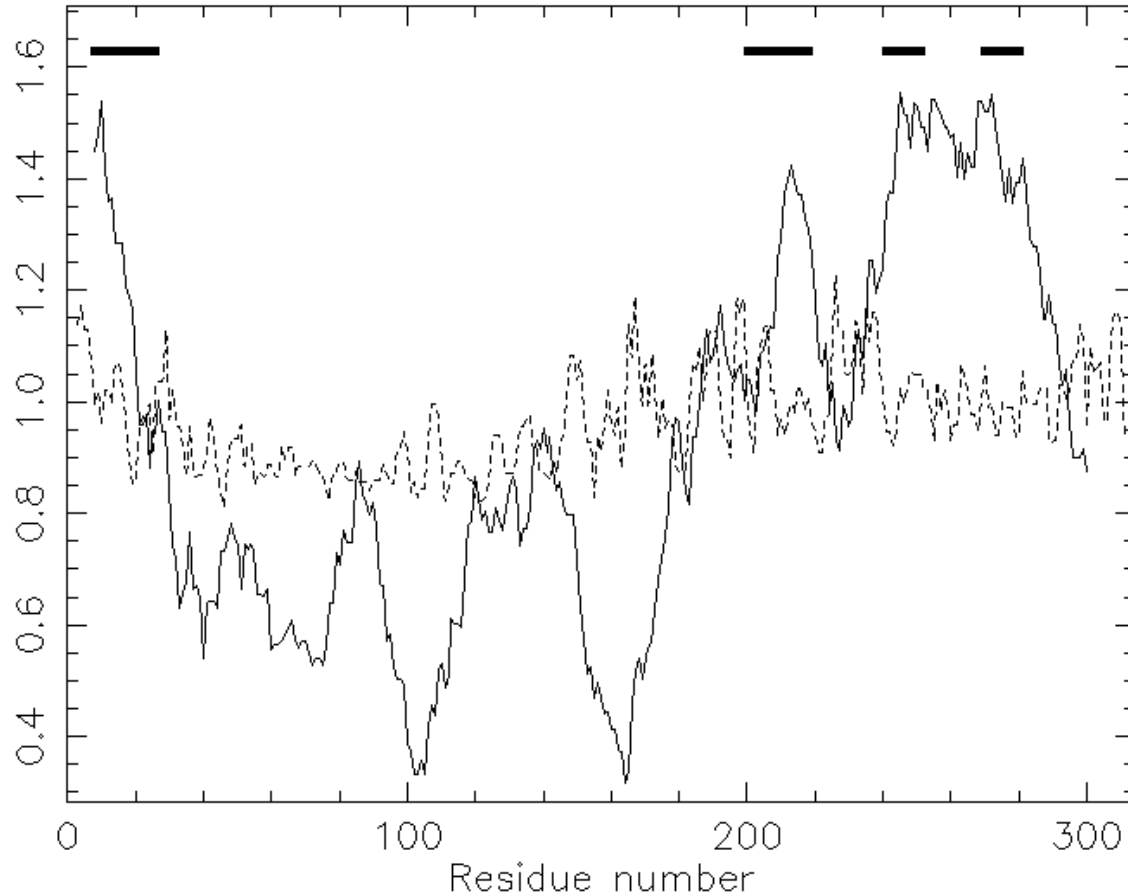
用garnier分析其二级结构， $\alpha$ 螺旋占34.9%， $\beta$ 折叠占18.1%， $\beta$ 转角占32.6%，无规则卷曲占19.8%。





# 蛋白质跨膜区分析

Tmap

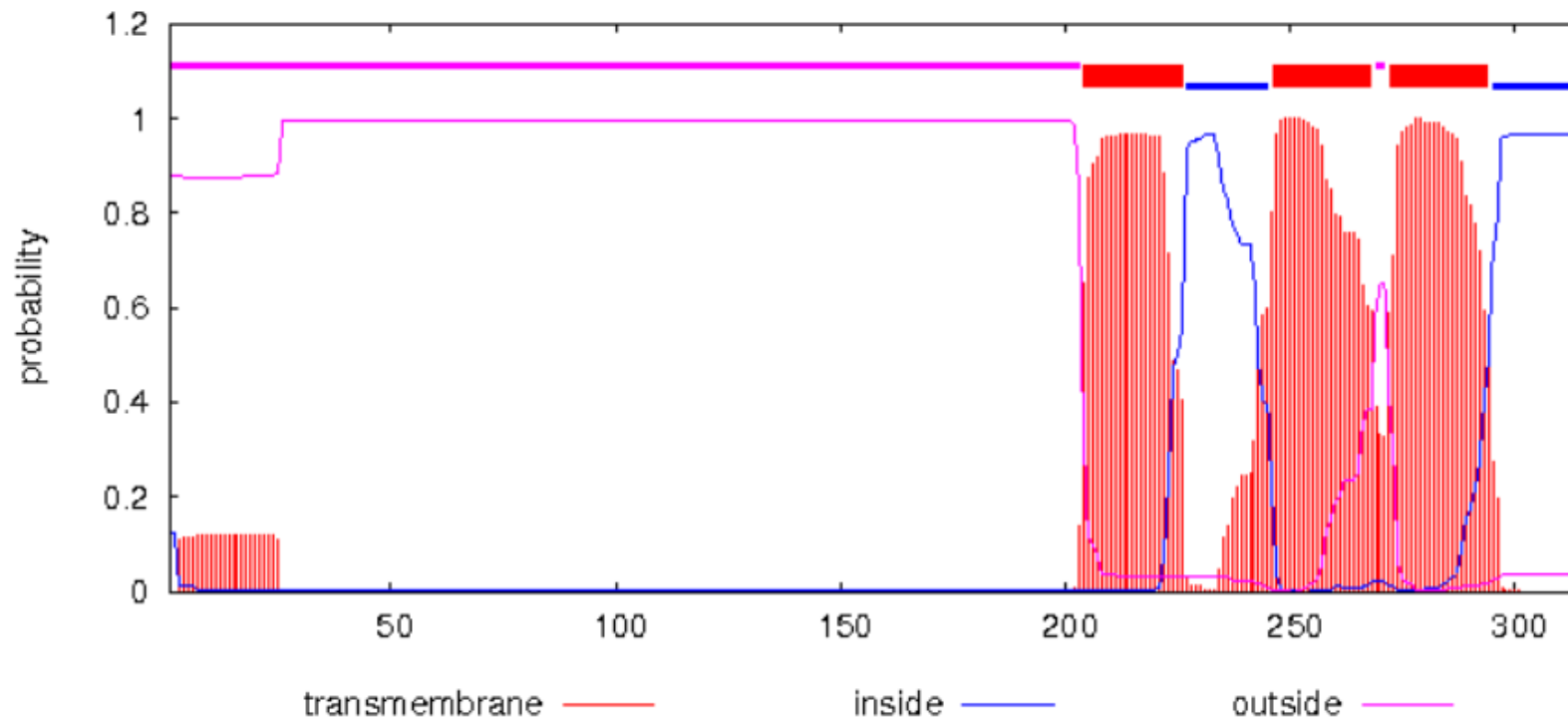


用Tmap预测其跨膜蛋白，有四个跨膜区。

Start	End	TransMem	Sequence
3	31	1	FFYLFVLFPIILLKFCECGPFLPLDRQLNP
195	223	2	FDWWLCDKPWQYALTLALFGFSLLSPCL
236	256	3	RSFIFDCFLTHLFLFLIAFCA
265	285	4	LVVMTFNVGVFFAVITGYTVG



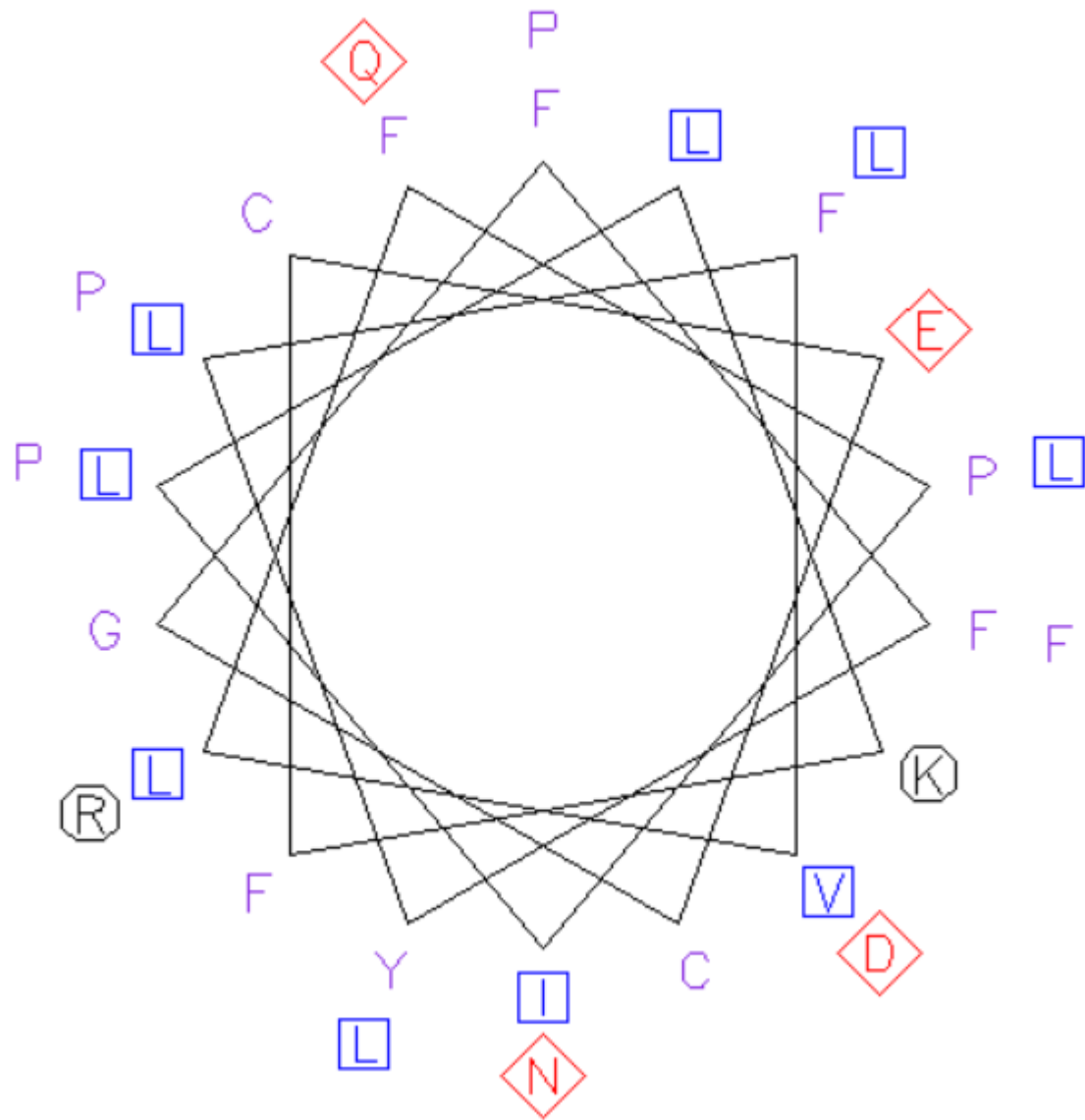
# TMHMM posterior probabilities for WEBSEQUENCE

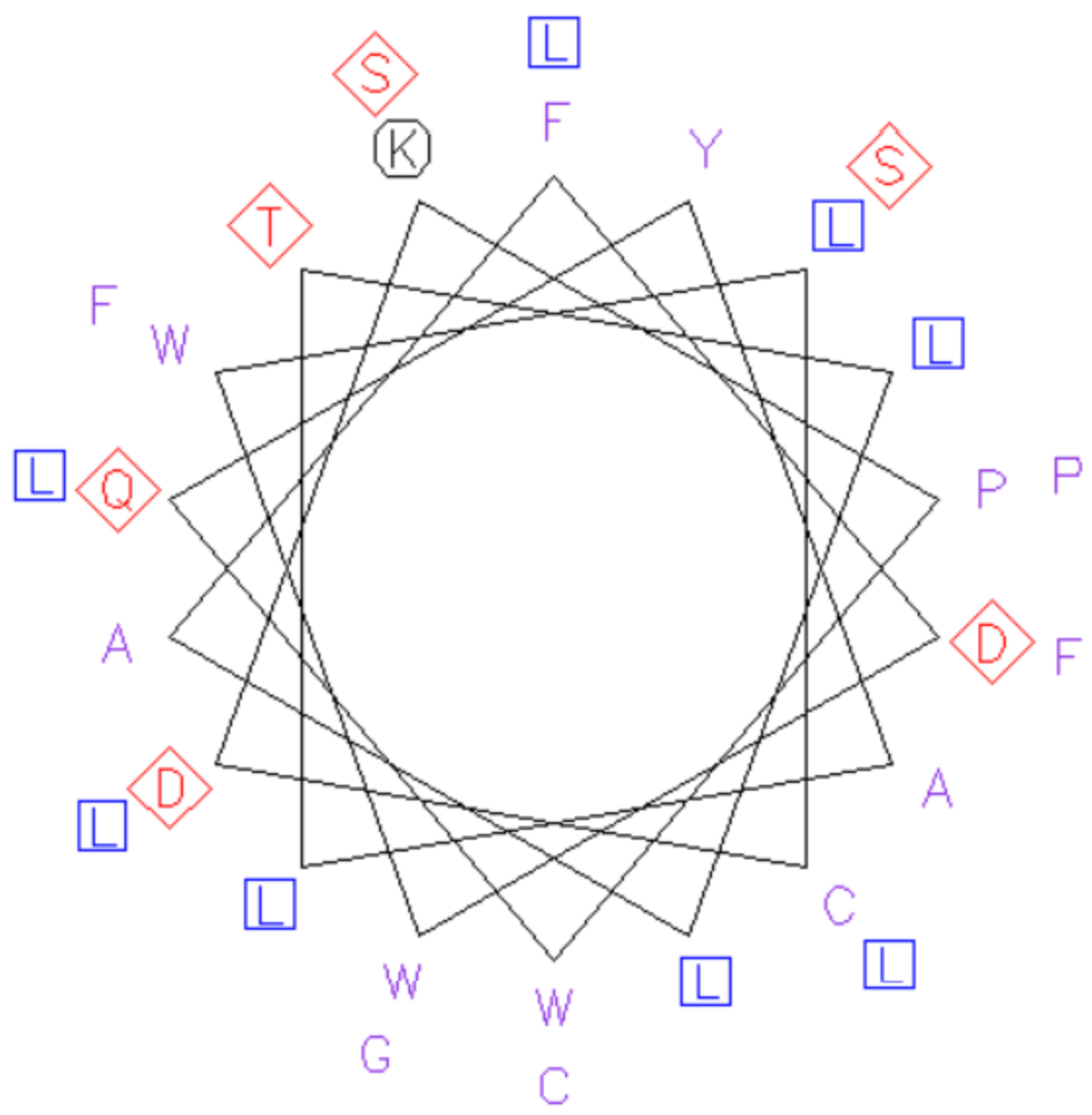


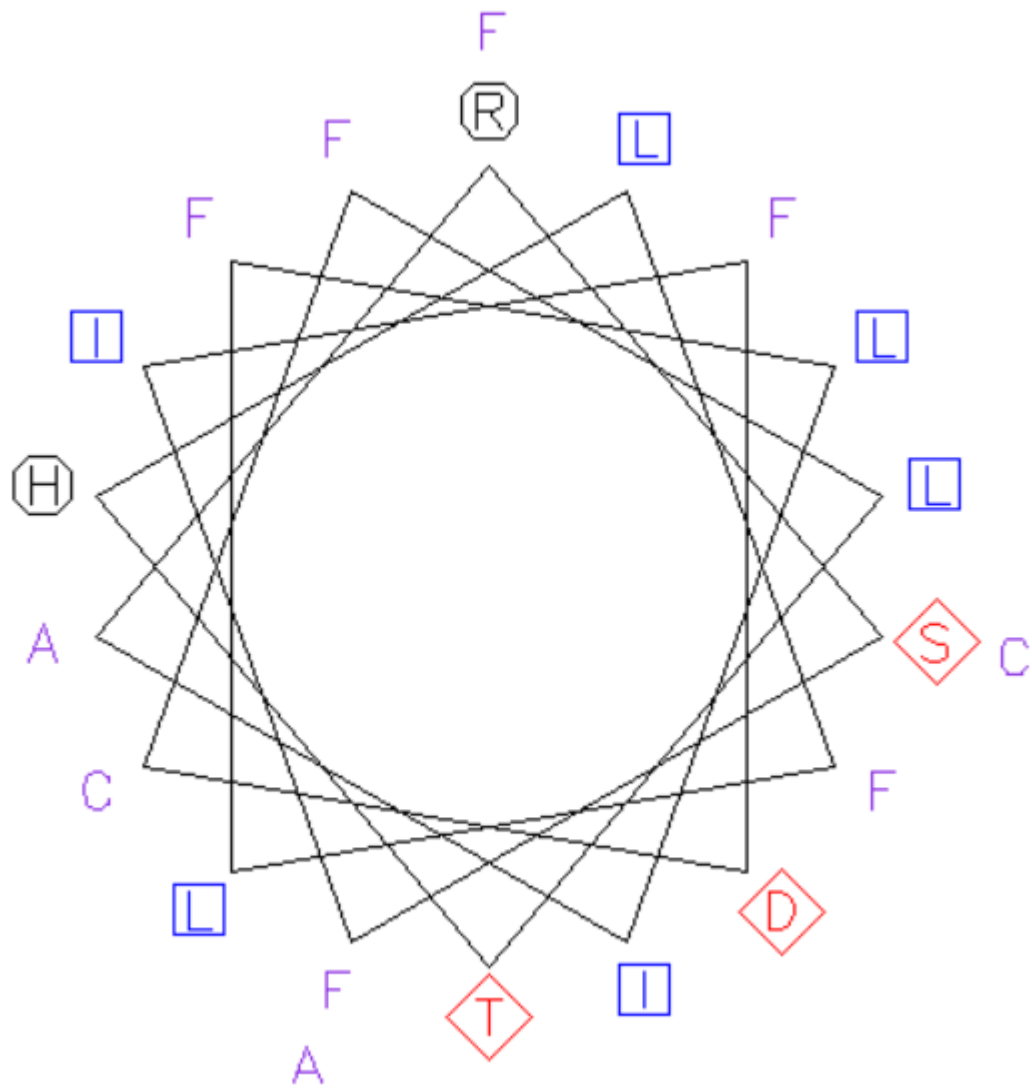
```

# WEBSEQUENCE Length: 314
# WEBSEQUENCE Number of predicted TMHs: 3
# WEBSEQUENCE Exp number of AAs in TMHs: 67.72717
# WEBSEQUENCE Exp number, first 60 AAs: 2.7078
# WEBSEQUENCE Total prob of N-in: 0.12232
WEBSEQUENCE TMHMM1.0 outside 1 203
WEBSEQUENCE TMHMM1.0 TMhelix 204 226
WEBSEQUENCE TMHMM1.0 inside 227 245
WEBSEQUENCE TMHMM1.0 TMhelix 246 268
WEBSEQUENCE TMHMM1.0 outside 269 271
WEBSEQUENCE TMHMM1.0 TMhelix 272 294
WEBSEQUENCE TMHMM1.0 inside 295 314
    
```

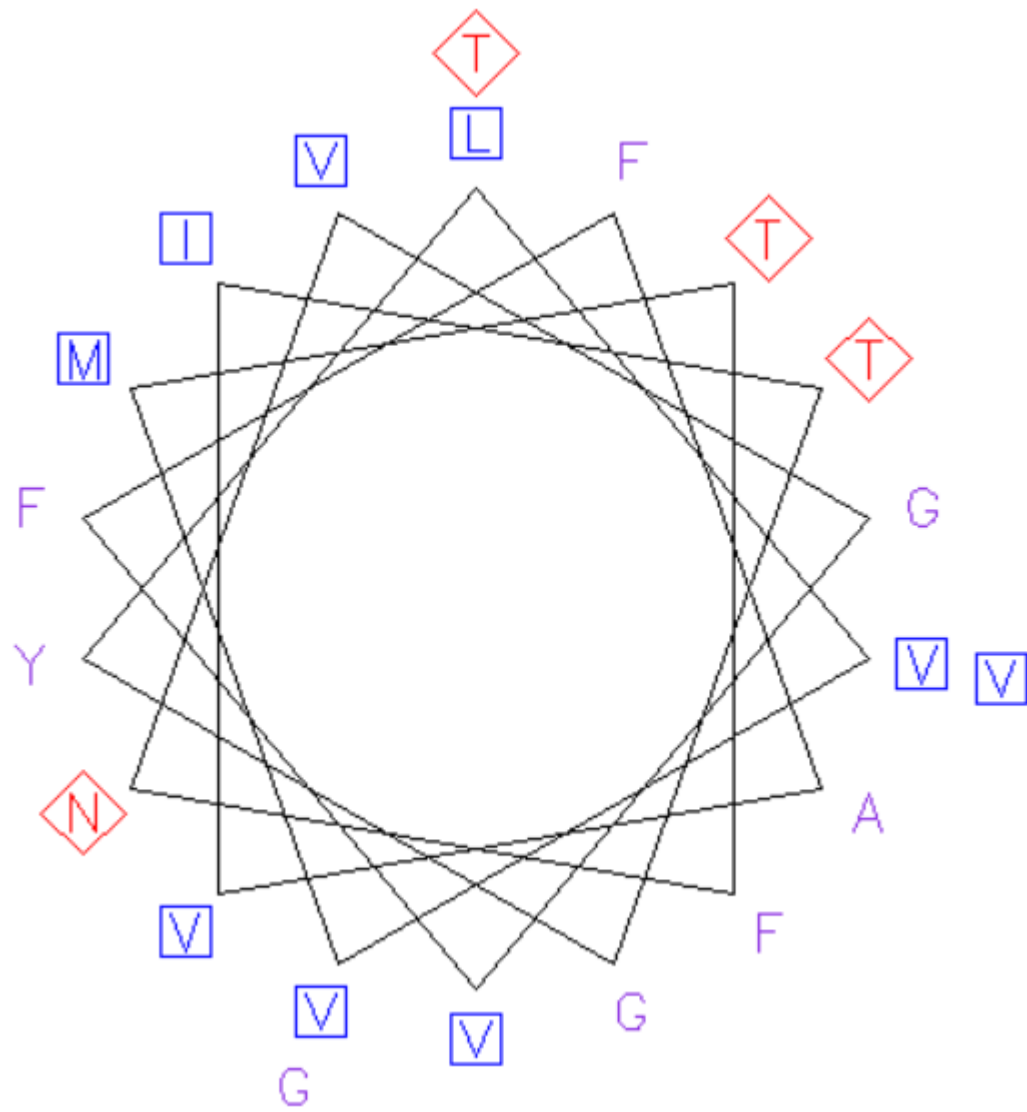




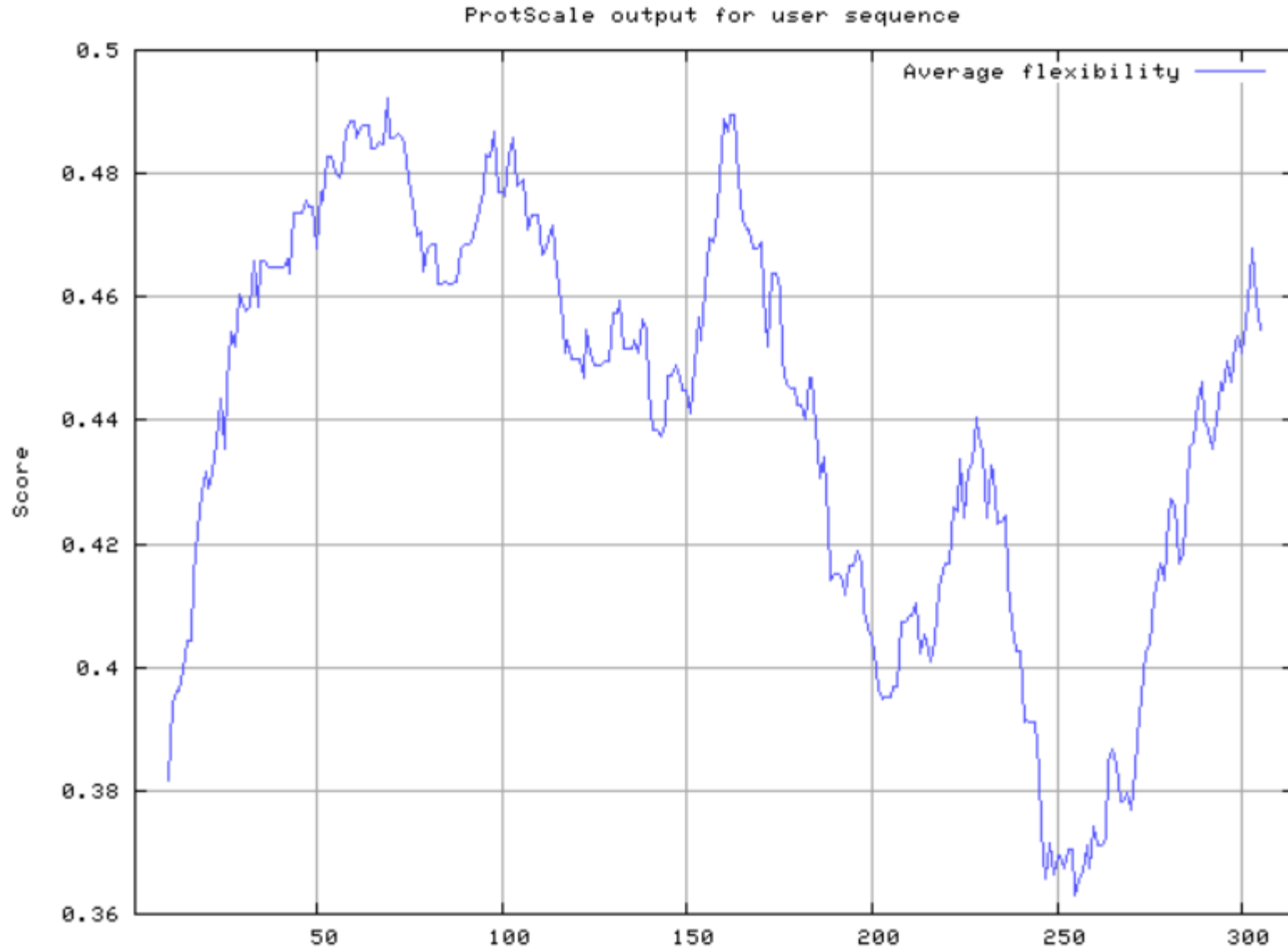








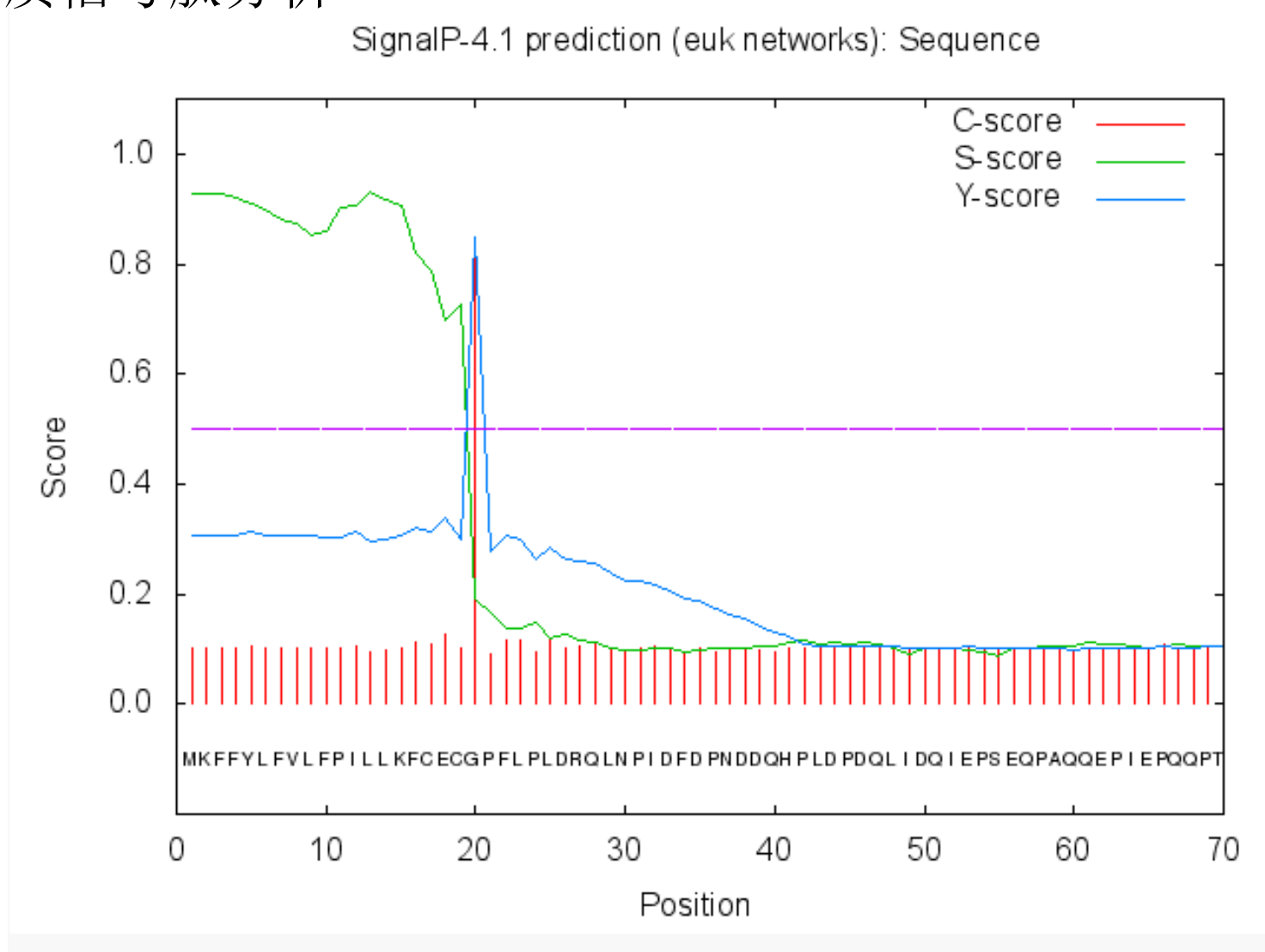
# 蛋白质柔韧性分析



用ProtScale预测柔韧性



# 蛋白质信号肽分析



SignalP 4.1 Server 分析其信号肽序列，位于1-19位氨基酸。

# 蛋白质亚细胞定位分析

```
### targetp v1.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions included.  
Using NON-PLANT networks.
```

Name	Len	mTP	SP	other	Loc	RC	TPlen
Sequence	314	0.018	0.963	0.069	S	1	19
cutoff		0.000	0.000	0.000			

用TargetP 1.1 Server 对其分析，显示其属于分泌型蛋白。



# 总结

- 该基因全长945bp，编码314个氨基酸。
- 分子量为35449.97 ku，等电点为4.4314。
- $\alpha$ 螺旋占34.9%， $\beta$ 折叠占18.1%， $\beta$ 转角占32.6%，无规则卷曲占19.8%。
- 分析其信号肽序列，位于1-19位氨基酸。
- 亚细胞定位，显示其属于分泌型蛋白。
- 3-4个跨膜区。



## 讨论

- Leu/Thr/Pro/Asp/His/Lys出现频率高，抗原指数就越大，形成抗原表位可能性也越高。
- 局部具亲水性，大多数高抗原指数区域与亲水区重合。（位于50—200位氨基酸）
- 序列中存在 $\beta$ 转角， $\beta$ 转角常作为其识别位点。（13—45位氨基酸和170—200位氨基酸处较多）
- 序列中带电荷氨基酸越多越有利于形成抗原决定簇。



综合氨基酸组成、亲水性、疏水性、  
二级结构、电荷、柔韧性等分析，  
预测该蛋白具潜在抗原性。

