

日本血吸虫抑制素 生物信息学分析

G15:

马帅 夏天奇

李莎 塔娜

报告内容：

① 研究背景

② 编码区核酸序列分析

③ 蛋白序列分析

④ 总结与致谢

研究背景

- **东方田鼠**是迄今由我国学者在血吸虫病疫区发现的唯一一种感染日本血吸虫后不致病的哺乳动物。因此东方田鼠是日本血吸虫的**抗性宿主**，东方田鼠血清特异识别的蛋白可能会对虫体的生长发育有影响。
- 经过筛选，**抑制素**能被东方田鼠的血清特异识别，因此对其进行生物信息学分析具有重要实用意义。

核酸序列分析

利用RT-PCR方法，克隆出的日本血吸虫抑制素编码区序列及注释信息如下：

```
LOCUS       FN318529                1003 bp    mRNA     linear   INV 03-SEP-2009
DEFINITION  Schistosoma japonicum isolate Anhui full length mRNA clone
            SJFCE2699.002|FSE001-P00026-L10, complete sequence.
ACCESSION   FN318529
VERSION     FN318529.1  GI:226484697
DBLINK      BioProject: PRJEA34885
KEYWORDS    FLI_CDNA; cap trapper.
SOURCE      Schistosoma japonicum
  ORGANISM  Schistosoma japonicum
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
```

核酸序列分析

---密码子使用

用WebLab中cusp(v6.0.1) 软件分析编码区的密码子使用:

```
#CdsCount: 1

#Coding GC 42.18%
#1st letter GC 58.91%
#2nd letter GC 39.64%
#3rd letter GC 28.00%

#Codon AA Fraction Frequency Number
GCA  A   0.240  21.818    6
GCC  A   0.240  21.818    6
GCG  A   0.080   7.273    2
GCT  A   0.440  40.000   11
TGC  C   0.000   0.000    0
TGT  C   0.000   0.000    0
GAC  D   0.333  14.545    4
GAT  D   0.667  29.091    8
GAA  E   0.739  61.818   17
GAG  E   0.261  21.818    6
TTC  F   0.417  18.182    5
TTT  F   0.583  25.455    7
GGA  G   0.056   3.636    1
GGC  G   0.000   0.000    0
GGG  G   0.111   7.273    2
GGT  G   0.833  54.545   15
```

CAC	H	0.500	7.273	2
CAT	H	0.500	7.273	2
ATA	I	0.318	25.455	7
ATC	I	0.182	14.545	4
ATT	I	0.500	40.000	11
AAA	K	0.818	32.727	9
AAG	K	0.182	7.273	2
CTA	L	0.133	14.545	4
CTC	L	0.033	3.636	1
CTG	L	0.167	18.182	5
CTT	L	0.333	36.364	10
TTA	L	0.133	14.545	4
TTG	L	0.200	21.818	6
ATG	M	1.000	7.273	2
AAC	N	0.125	3.636	1
AAT	N	0.875	25.455	7
CCA	P	0.273	10.909	3
CCC	P	0.000	0.000	0
CCG	P	0.091	3.636	1
CCT	P	0.636	25.455	7
CAA	Q	0.714	36.364	10
CAG	Q	0.286	14.545	4
AGA	R	0.263	18.182	5
AGG	R	0.000	0.000	0
CGA	R	0.368	25.455	7
CGC	R	0.105	7.273	2
CGG	R	0.105	7.273	2
CGT	R	0.158	10.909	3
AGC	S	0.048	3.636	1
AGT	S	0.143	10.909	3

TCA	S	0.286	21.818	6
TCC	S	0.095	7.273	2
TCG	S	0.190	14.545	4
TCT	S	0.238	18.182	5
ACA	T	0.286	14.545	4
ACC	T	0.143	7.273	2
ACG	T	0.143	7.273	2
ACT	T	0.429	21.818	6
GTA	V	0.238	18.182	5
GTC	V	0.143	10.909	3
GTG	V	0.190	14.545	4
GTT	V	0.429	32.727	9
TGG	W	1.000	3.636	1
TAC	Y	0.167	3.636	1
TAT	Y	0.833	18.182	5
TAA	*	1.000	3.636	1
TAG	*	0.000	0.000	0
TGA	*	0.000	0.000	0

核酸序列分析

---限制位点分析

- 用**restrict(v6.0.1)**软件分析该基因序列中的限制性核酸内切酶位点，分别把限制位点的长度限制为**4、6、和8 bp**时，其中限制位点的个数则为**698个、130个、23个**。我们对这个基因序列进行操作时要考虑这些限制位点，**以防止破坏读码框；也便于根据实验室常用酶使用特定限制位点。**

蛋白质序列分析

---理化性质

利用WebLab中pepstats(v6.0.1)和ExPASy网站ProtParam软件分析

Total number of negatively charged residues (Asp + Glu): 35
Total number of positively charged residues (Arg + Lys): 30

Atomic composition:

Carbon	C	1354
Hydrogen	H	2181
Nitrogen	N	373
Oxygen	O	408
Sulfur	S	2

Number of amino acids: 274

Molecular weight: 30277.5

Formula: C₁₃₅₄H₂₁₈₁N₃₇₃O₄₀₈S₂
Total number of atoms: 4318

Theoretical pI: 5.54

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

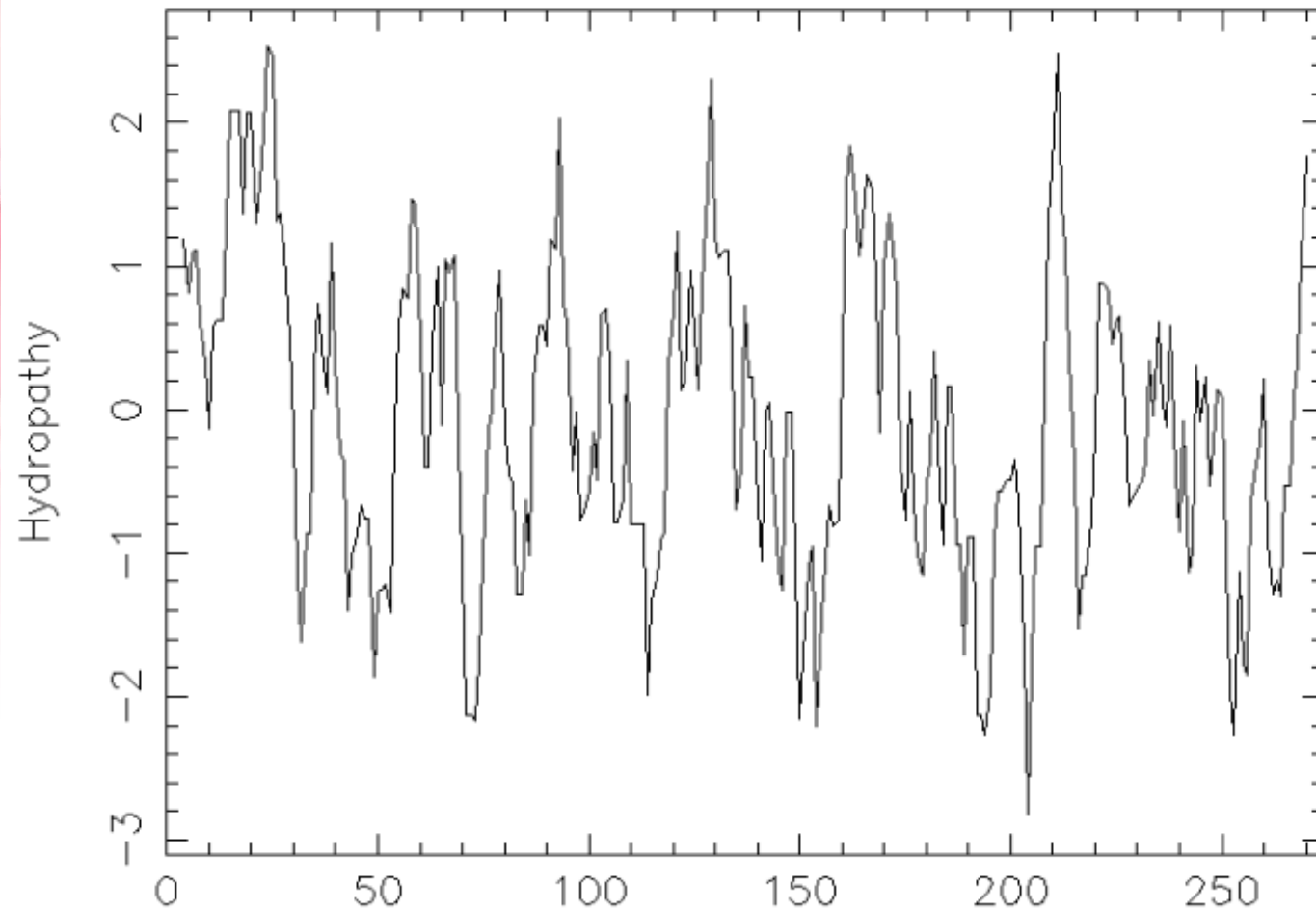
Ext. coefficient	14440
Abs 0.1% (=1 g/l)	0.477

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	78	28.467
Small	(A+B+C+D+G+N+P+S+T+V)	130	47.445
Aliphatic	(A+I+L+V)	98	35.766
Aromatic	(F+H+W+Y)	23	8.394
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	148	54.015
Polar	(D+E+H+K+N+Q+R+S+T+Z)	126	45.985
Charged	(B+D+E+H+K+R+Z)	69	25.182
Basic	(H+K+R)	34	12.409
Acidic	(B+D+E+Z)	35	12.774

蛋白序列分析

- 利用pepwindow(v6.0.1) 软件对抑制素进行疏水性亲水性分析

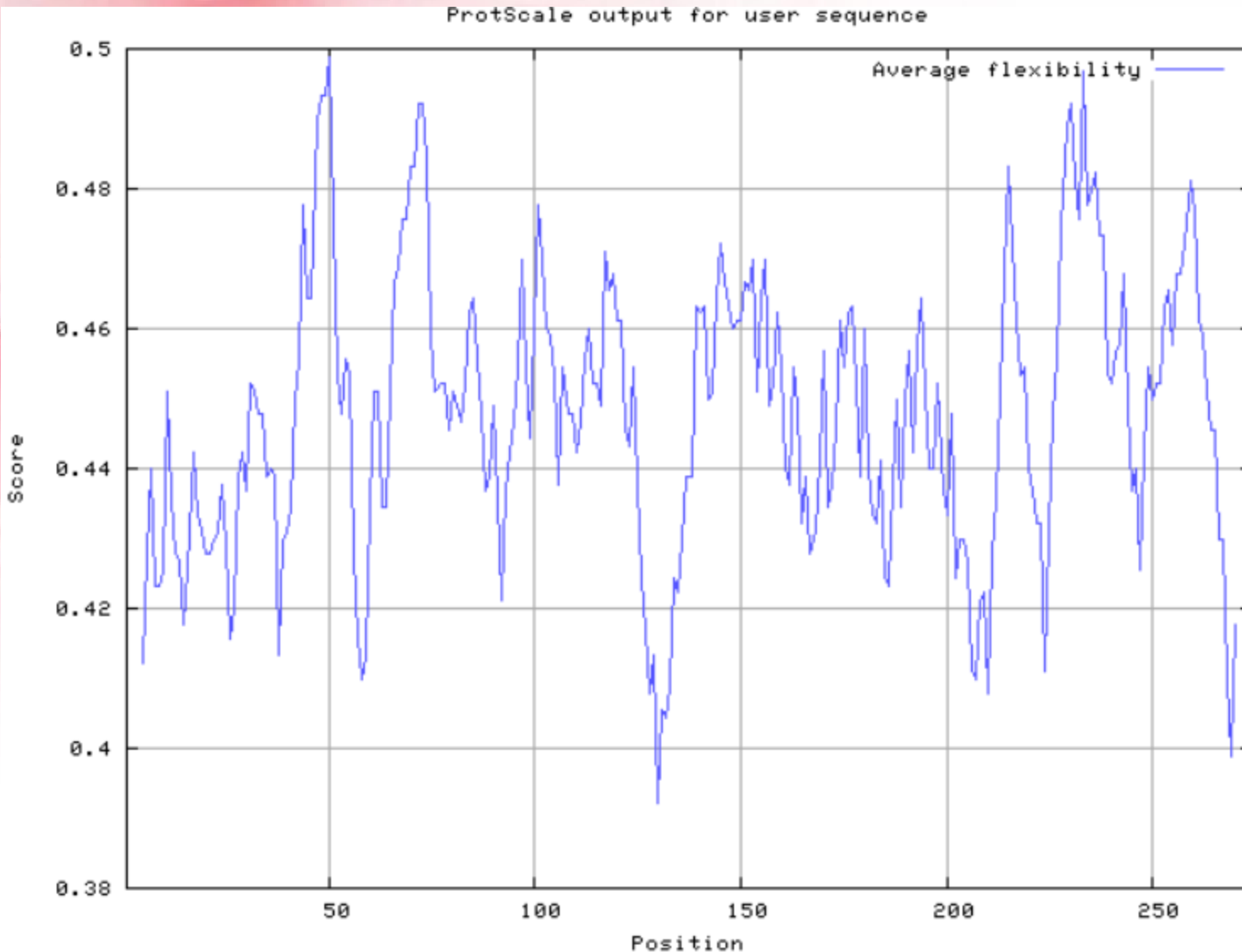
Kyte-Doolittle Plot



氨基酸分值越**低**亲水性**强**，分值越**高**疏水性**越强**。此蛋白序列亲水性得分比较均，且亲水疏水分布均匀。在机体内，疏水性残基一般埋在蛋白质内部，而**亲水性残基位于表面**，蛋白质的**亲水部位与蛋白抗原表位**也有密切的联系。

蛋白序列分析

- 利用ProtScale对其进行柔性刚性分析



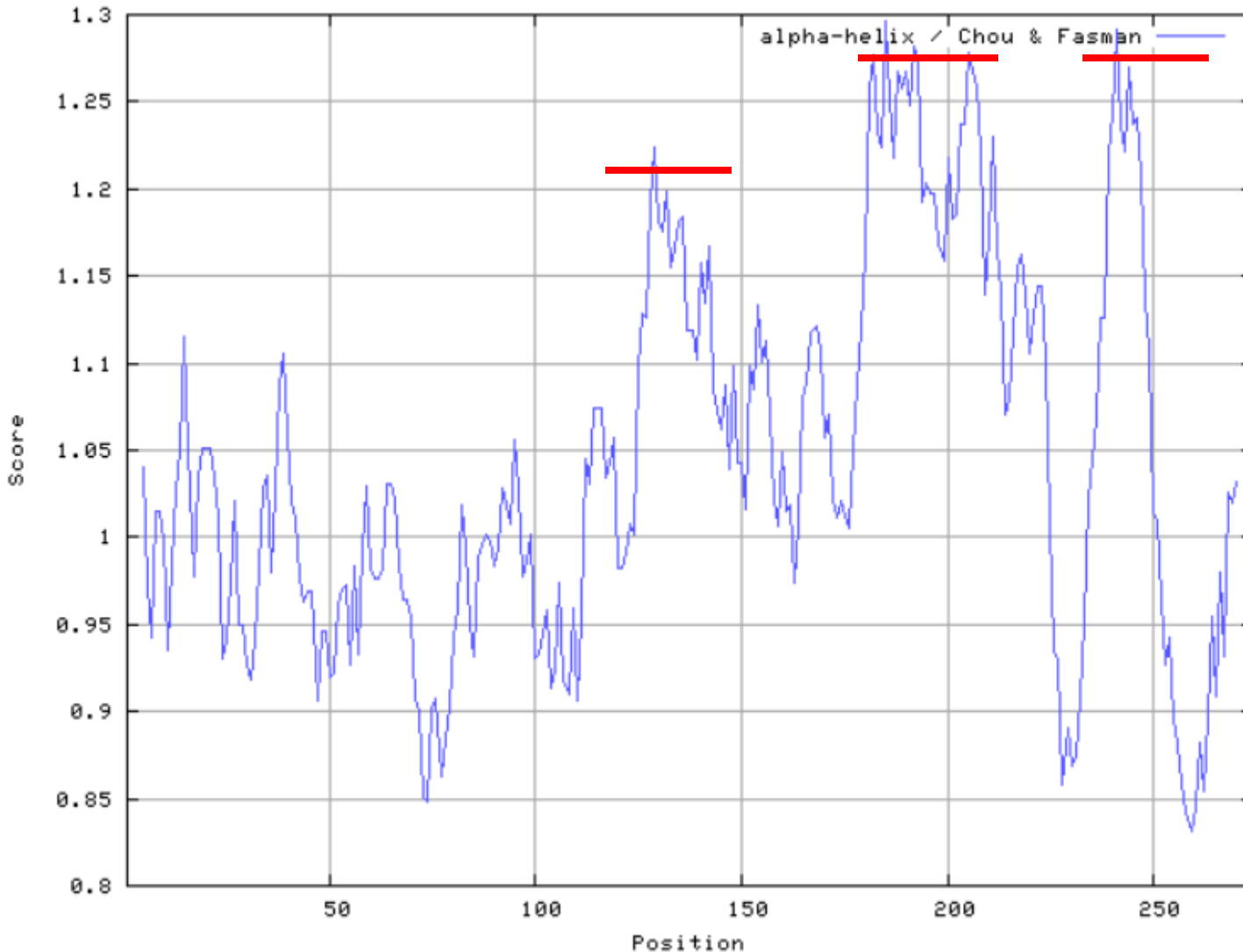
有3个柔性较大的区域，可能是高级结构中的不规则卷曲部分。

一定的柔韧性有利于与蛋白抗体结合，活性强的氨基酸残基即柔韧性大的位点，容易形成抗原表位，

蛋白质序列分析

- 利用ProtScale对其进行 α 螺旋预测

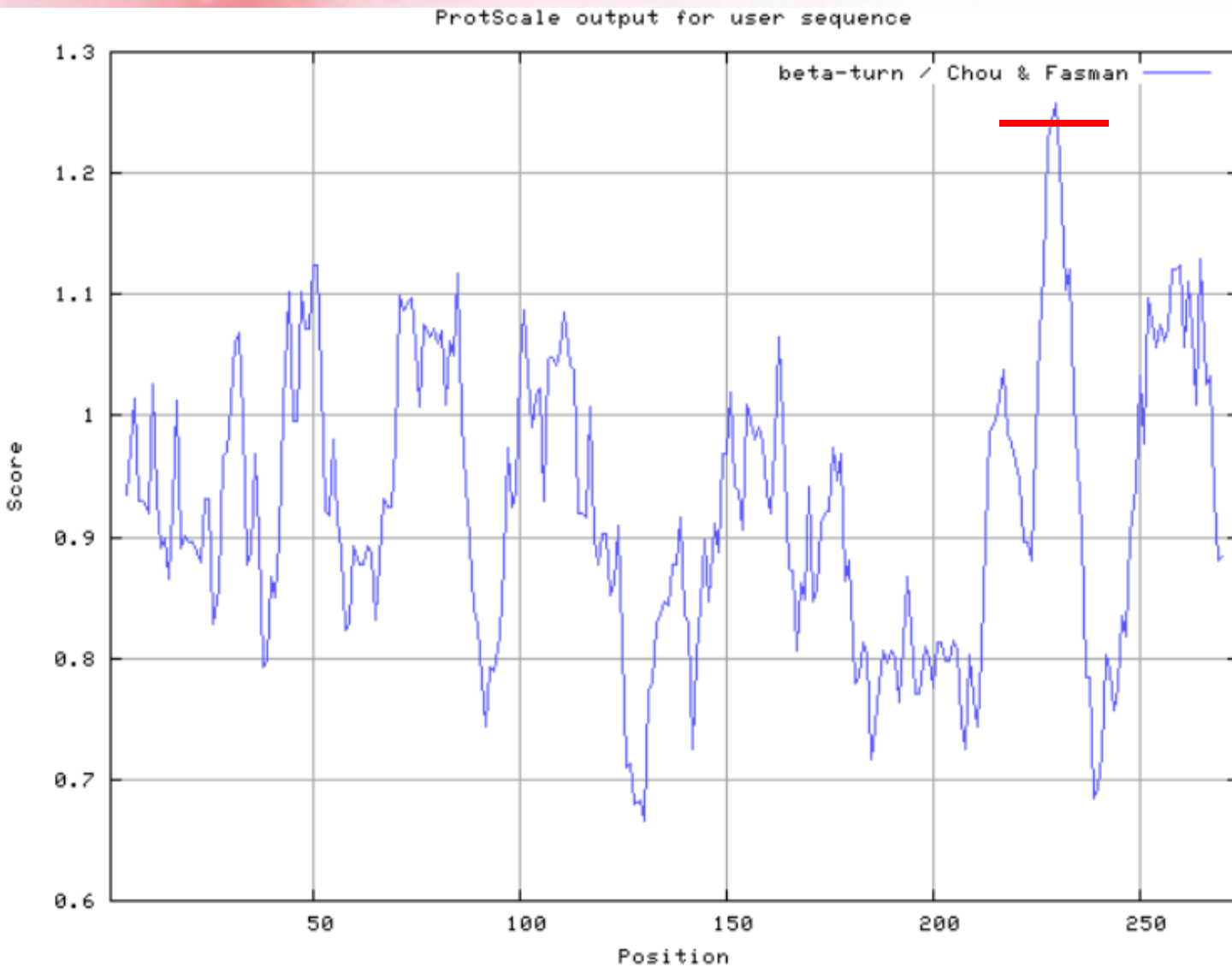
ProtScale output for user sequence



有3个明显的 α 螺旋区域

蛋白质序列分析

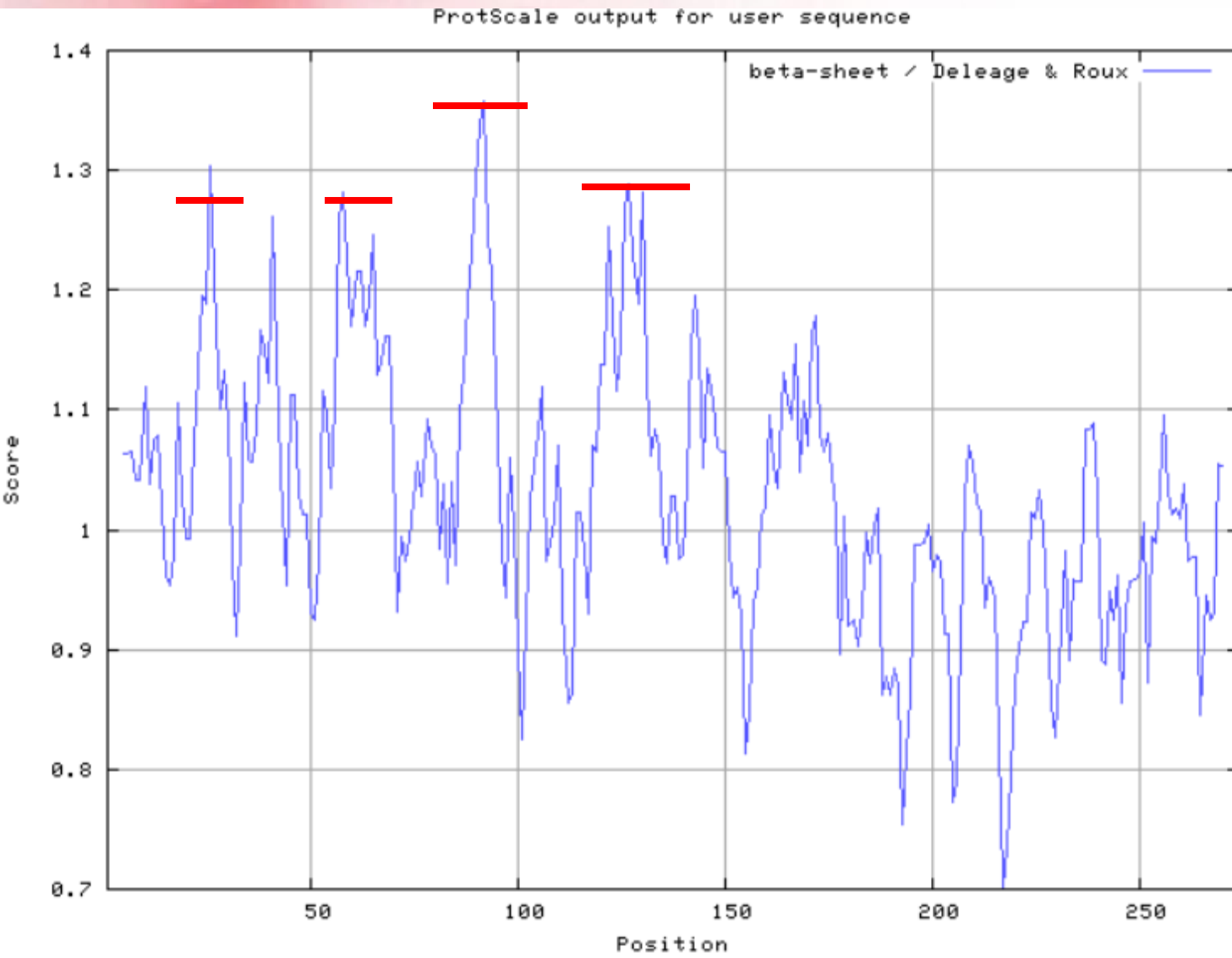
- β 转角预测



有一个可能的
 β 转角
区域

蛋白质序列分析

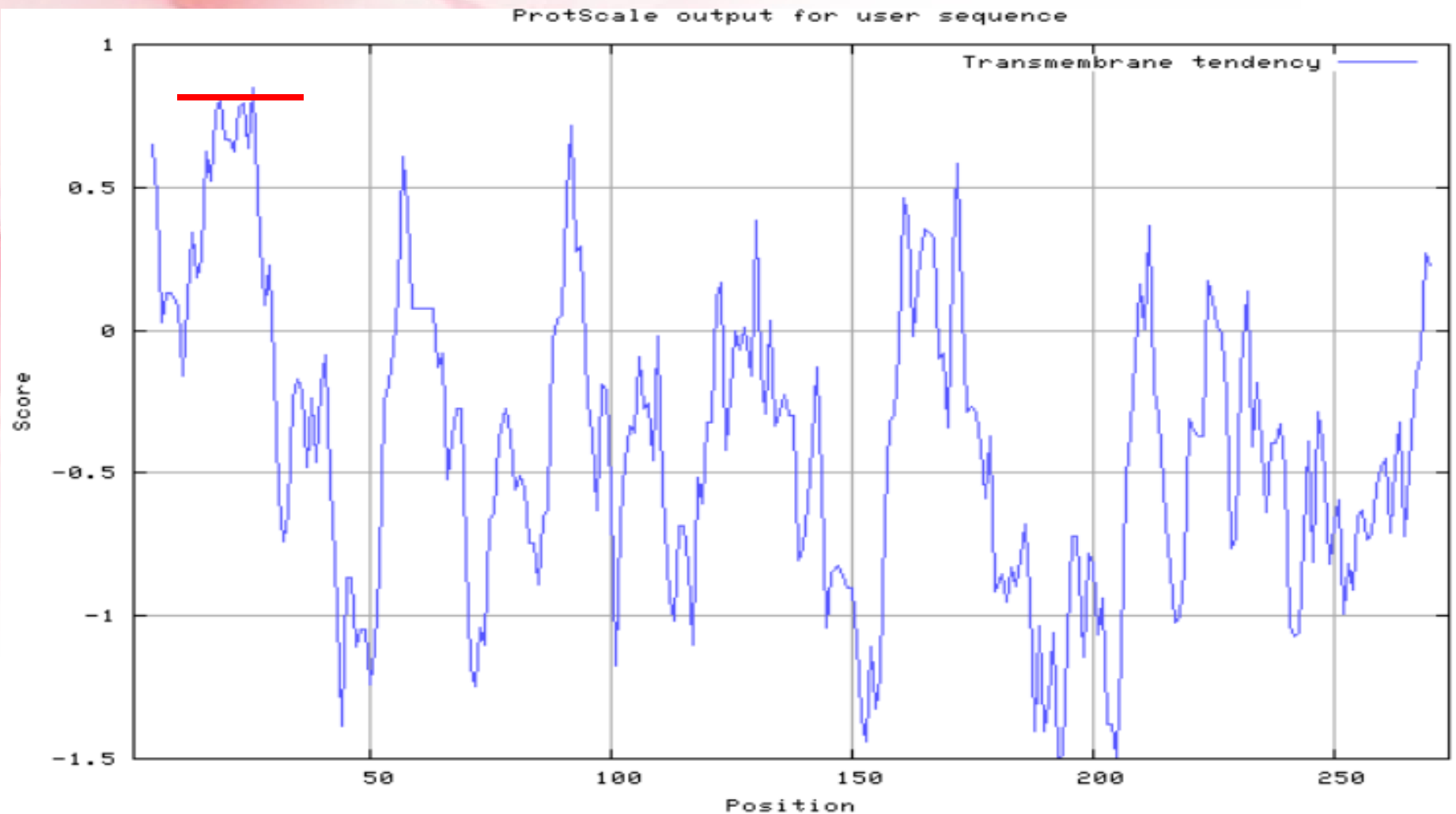
- β 折叠预测



存在4个可能的 β 折叠区域

蛋白质序列分析

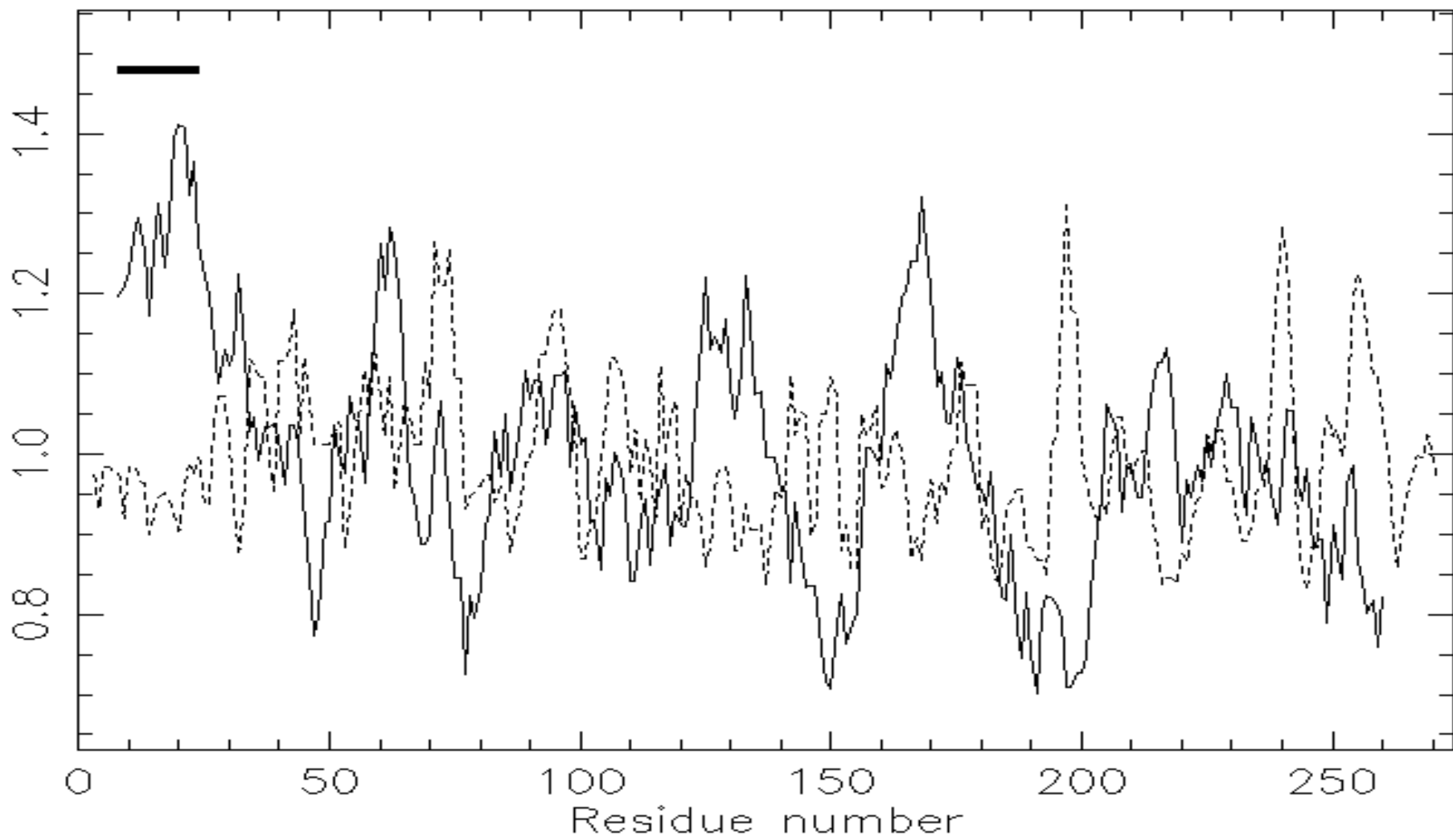
- 预测蛋白质序列中可能的跨膜螺旋
(1) 用ExPASy, ProtScale



蛋白质序列分析

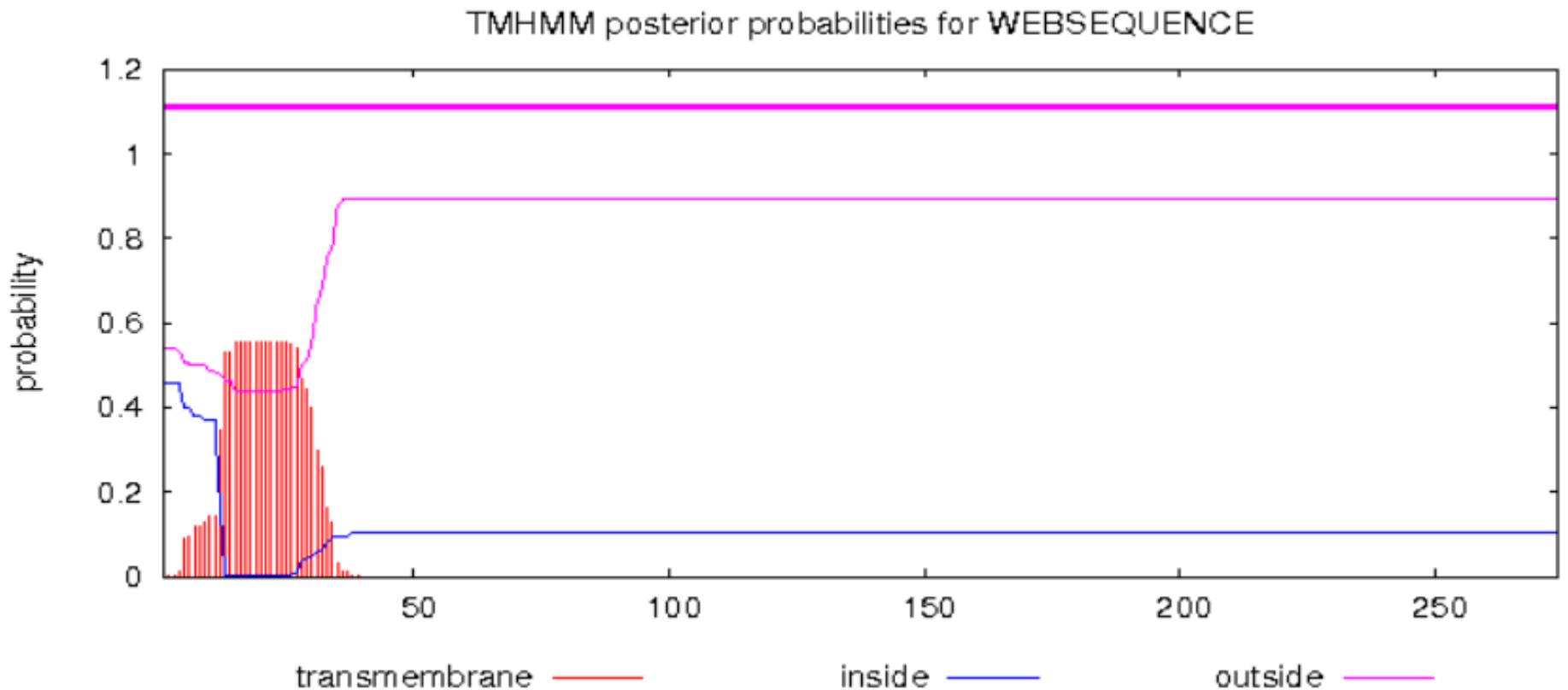
(2) 用webleb中Tmap软件程序

Tmap



蛋白质序列分析

(3) TMHMM



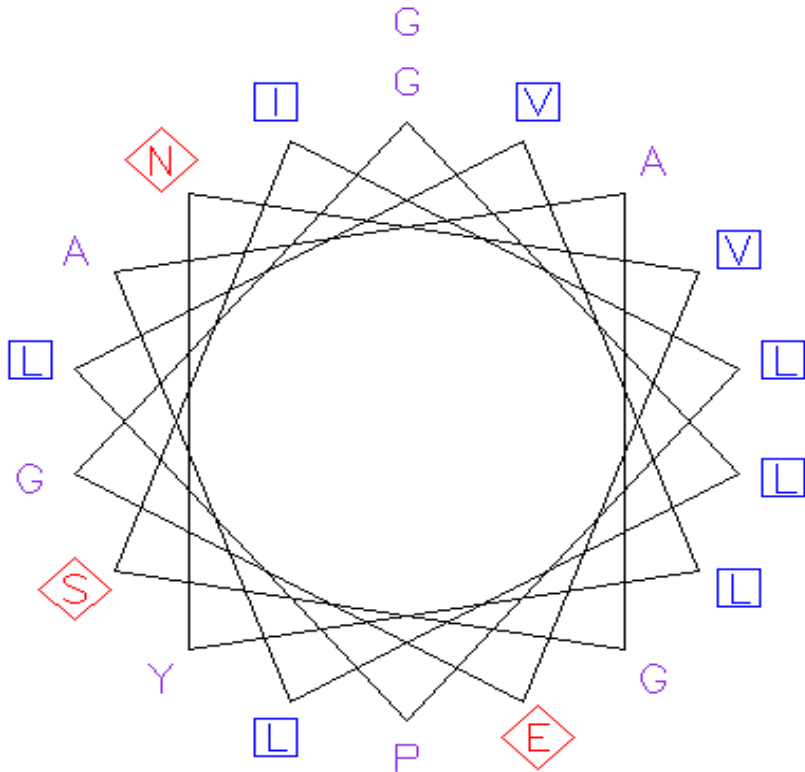
以上三个软件预测的跨膜螺旋结果基本一致，存在一个跨膜结构域。位于14-32。

蛋白序列分析

- 蛋白跨膜螺旋的pepwheel预测 (14-32)
GLLAAGSILPLVLYNVEGG

Helical wheel of raw::635830

Thu 20 Jun 2013 11:47:27

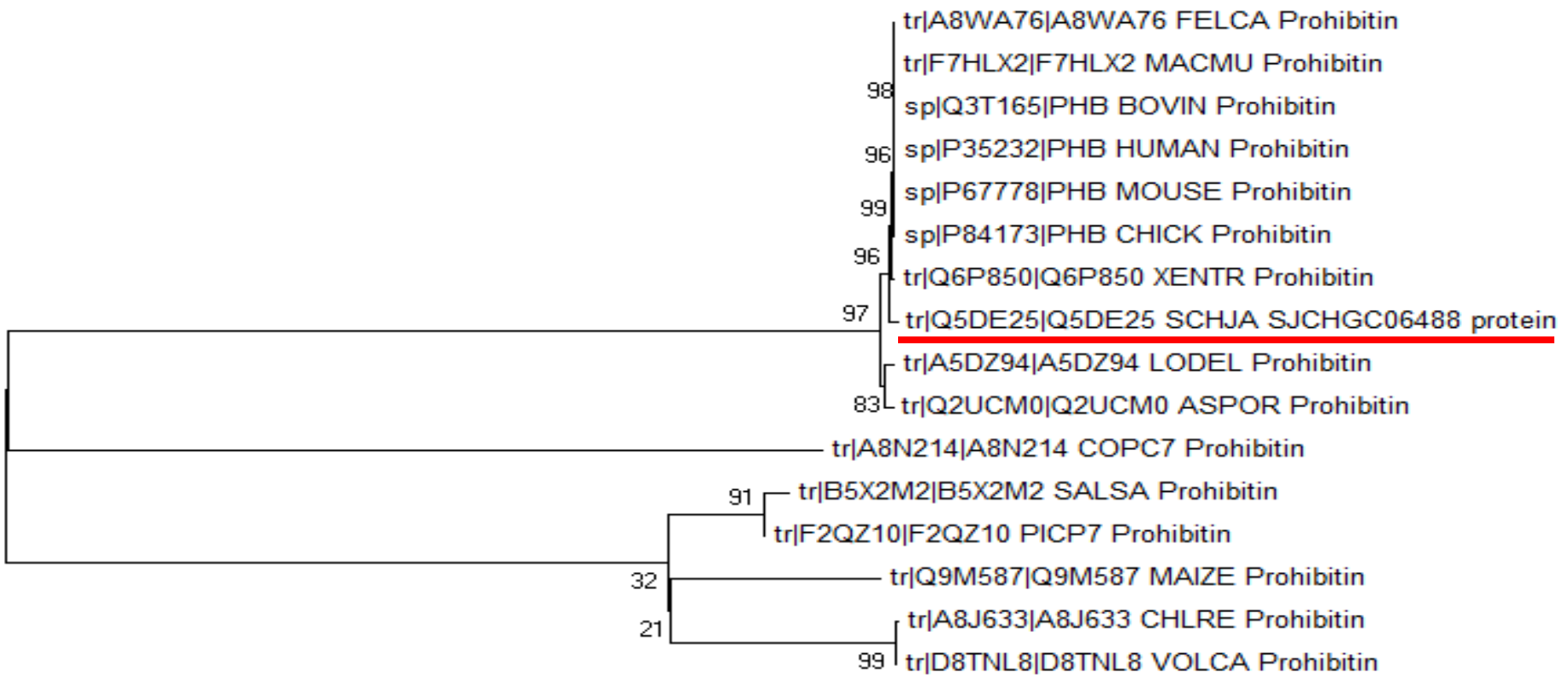


轮状图的周围是螺旋上的氨基酸，紫色字母不带框的是非极性氨基酸，具有疏水性；蓝色的带方框的是非极性氨基酸，疏水性比紫色的稍强，红色方框的是亲水性氨基酸。总体来看，所预测的区段几乎有疏水性氨基酸组成。

蛋白序列分析

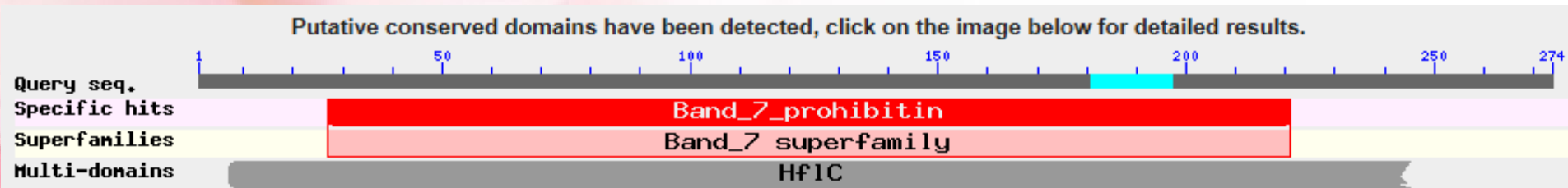
---进化树

- 从UniProt数据库中下载抑制素的蛋白序列，共下载16条序列，并选取其中的由270-285个氨基酸组成的片段，用MEGA5.05软件构建进化树，结果显示如图



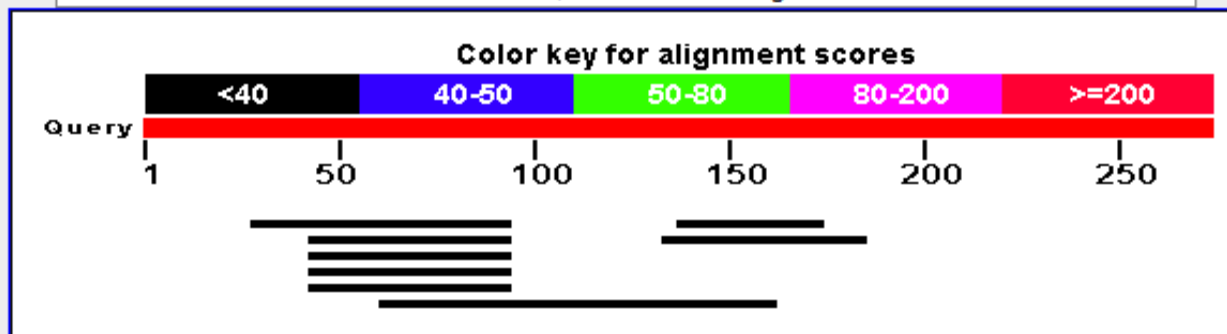
蛋白质序列分析

在同源建模之前，先用BLAST在pdb数据库中搜索，有没有相似性很高的模板可选择，如果相似性低于30%，则swiss-model是建不出来的，或者说即使建出来，准确率也很低，没有参考价值。



Distribution of 8 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



相似性太低，同源建模失败

总结

通过日本血吸虫抑制素蛋白进行的生物信息学分析，对这个蛋白有了**基本的了解**，同时**预测了其部分二级结构和性质**，**为后续试验提供了根本性的指导!**

致谢

- 感谢罗老师在这一学期中对我们的教育与指导！
- 感谢组内成员所作出的努力！

*Thanks for your
attention !!!*