

Bioinformatic Analysis of F Protein in PPRV

报告人：第10组

乔洋洋 刘静宜 刘伟 李君君

OUTLINE:

➤ **Background**

➤ ***Bioinformatic Analysis***

➤ **Reference**

Background

- **小反刍兽疫 (Peste des Petits Ruminants , PPR) 是由PPRV引起的小反刍兽一种严重传染病，特别是山羊极度易感。该病在世界上被认为是十分重要的烈性传染病之一，在我国该病也被列为一类动物传染病。**

- **PPRV属于副粘病毒科的麻疹病毒属成员。其基因组结构为单股负链RNA。目前，PPRV的许多毒株基因组序列已经测序完毕，其RNA链的3'至5'端依次分布着N-P-M-F-H-L,共六个基因，分别编码各自对应的6种结构蛋白和2种非结构蛋白。**

- **F蛋白全称是融合性糖蛋白
(Fusion glycoprotein)**
- **F基因在麻疹病毒属中高度保守**
- **其5' 端的序列具有病毒特异性**
- **PPRV的各种毒株就是根据F基因序列的遗传特性而分类的。**

研究目的：

本研究利用生物信息学的方法，对F蛋白质的基因及相应氨基酸序列的理化性质、结构特征、及系统演化关系等进行预测和分析，为深入开展F蛋白表达调控的分子机制的研究提供理论依据。

生物信息学分析

1. 在Uniprot中搜索，F蛋白的部分氨基酸序列其长度为106：

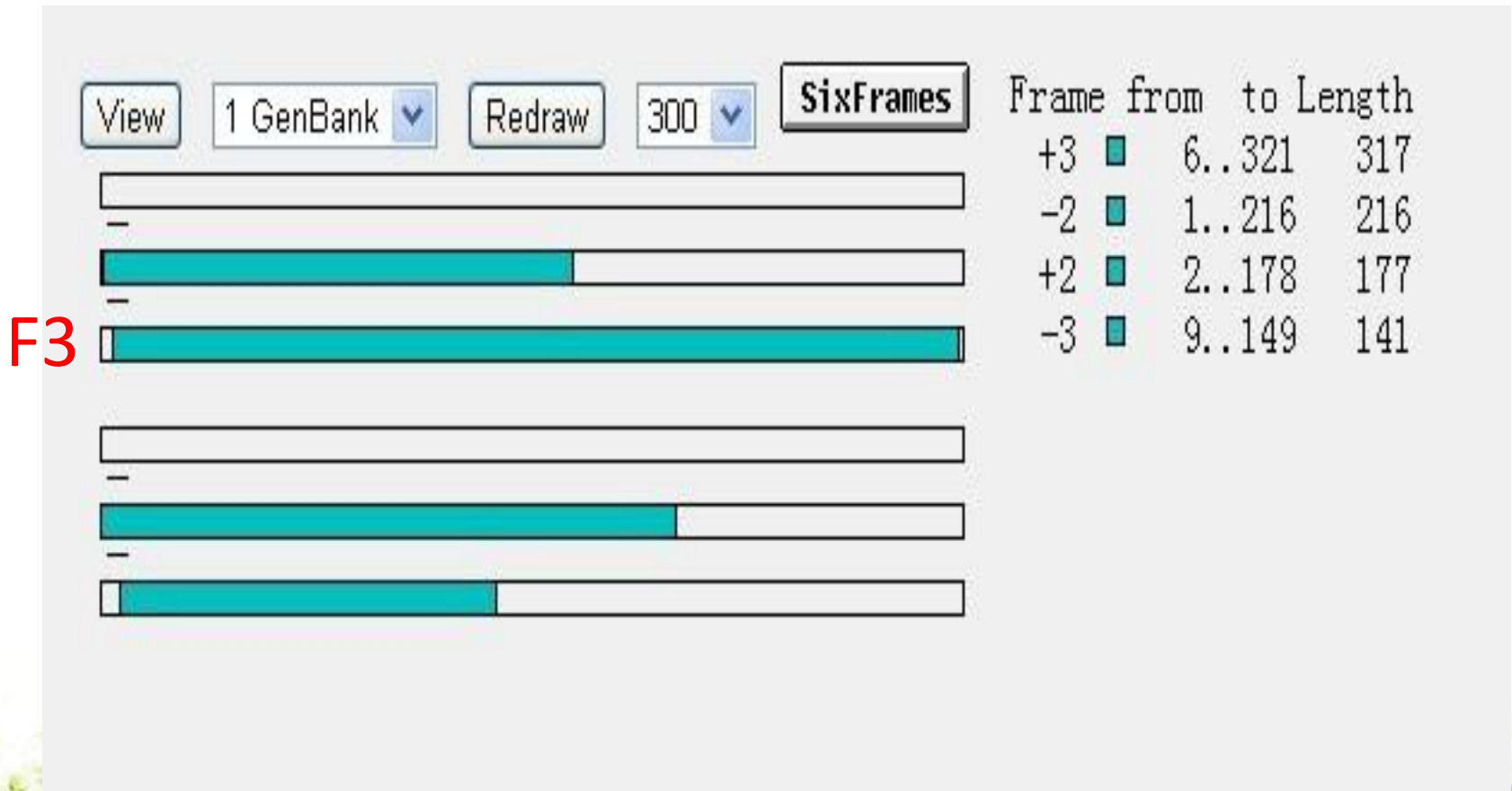
- **ALSVITKNVRPIQTLTPGRRTRRFAGAVLAG
VALGVATAAQITAGVALHQSLMNSQAIESL
KTSLEKSNQAIEEIRLANKETILAVQGVQDYI
NNELVPSVHRMSC**

2.F蛋白部分编码区序列：

- >gi|18766380|gb|AF464892.1| Peste-des-petits-ruminants virus clone Mdn96 fusion protein (F) gene, partialcds

**ATGCTCTGTCAGTGATAACCAAGAATGTAAGACCAAT
TCAAACCTCTGACACCTGGGCGTGGAACCTCGCCGTTTT
GCTGGAGCTGTTCTGGCCGGAGTAGCACTTGGAGTT
GCGACAGCCGCTCAGATAACTGCAGGAGTCGCCCTTC
ATCAATCATTGATGAACTCCCAAGCAATTGAGAGTTT
AAAACCCAGTCTTGAGAAGTCGAATCAGGCAATAGA
AGAAATCAGACTTGCAAATAAGGAGACCATACTGGC
AGTACAGGGCGTCCAGGATTATATCAACAATGAGCTT
GTCCCTTCTGTTCACAGAATGTCATGCGA**

3. ORF Finder (Open Reading Frame Finder)



4.用showorf软件分析F蛋白的mRNA

```
SHOWORF of from 1 to 322
-----|-----|-----|-----|
1 ATGCTCTGTCAGTGATAACCAAGAATGTAAGACCAATTCAAACCTCTGACA 50
F3 1 A L S V I T K N V R P I Q T L T 16
-----|-----|-----|-----|
51 CCTGGGCGTAGAACTCGCCGTTTTGCTGGAGCTGTTCTGGCCGGAGTAGC 100
F3 17 P G R R T R R F A G A V L A G V A 33
-----|-----|-----|-----|
101 ACTTGGAGTTGCGACAGCCGCTCAGATAACTGCAGGAGTCGCCCTTCATC 150
F3 34 L G V A T A A Q I T A G V A L H Q 50
-----|-----|-----|-----|
151 AATCATTGATGAACTCCCAAGCAATTGAGAGTTTAAAAACCAGTCTTGAG 200
F3 51 S L M N S Q A I E S L K T S L E 66
-----|-----|-----|-----|
201 AAGTCGAATCAGGCAATAGAAGAAATCAGACTTGCAAATAAGGAGACCAT 250
F3 67 K S N Q A I E E I R L A N K E T I 83
-----|-----|-----|-----|
251 ACTGGCAGTACAGGGCGTCCAGGATTATATCAACAATGAGCTTGTCCCCT 300
F3 84 L A V Q G V Q D Y I N N E L V P S 100
-----|-----|-----|
301 CTGTTTCATAGAATGTCATGCCA 322
F3 101 V H R M S C 106
```

与在Uniprot中搜索F蛋白的部分氨基酸序列相同

- 用WebLab中SixPack程序分析F蛋白全长mRNA序列读码框特征,发现只有以下序列有意义
- >_3_ORF1 Translation of in frame 3, ORF 1, threshold1, 106aa
- ALSVITKNVRPIQTLTPGRRTRRFAGAVLAGV
ALGVATAAQITAGVALHQSLMNSQAIESLKT
SLEKSNQAIEEIRLANKETILAVQGVQDYINN
ELVPSVHRMSCX



5.密码子使用特征

#CdsCount: 10+

+

#Coding GC 46.42%+

#1st letter GC 43.93%+

#2nd letter GC 37.38%+

#3rd letter GC 57.94%+

+

#Codon AA Fraction Frequency Number+

GCA	A	0.000	<u>0.000</u>	0+
GCC	A	0.333	9.346	1+
GCG	A	0.667	18.692	2+
GCT	A	0.000	<u>0.000</u>	0+
TGC	C	0.000	<u>0.000</u>	0+
TGT	C	1.000	18.692	2+
GAC	D	1.000	18.692	2+
GAT	D	0.000	<u>0.000</u>	0+

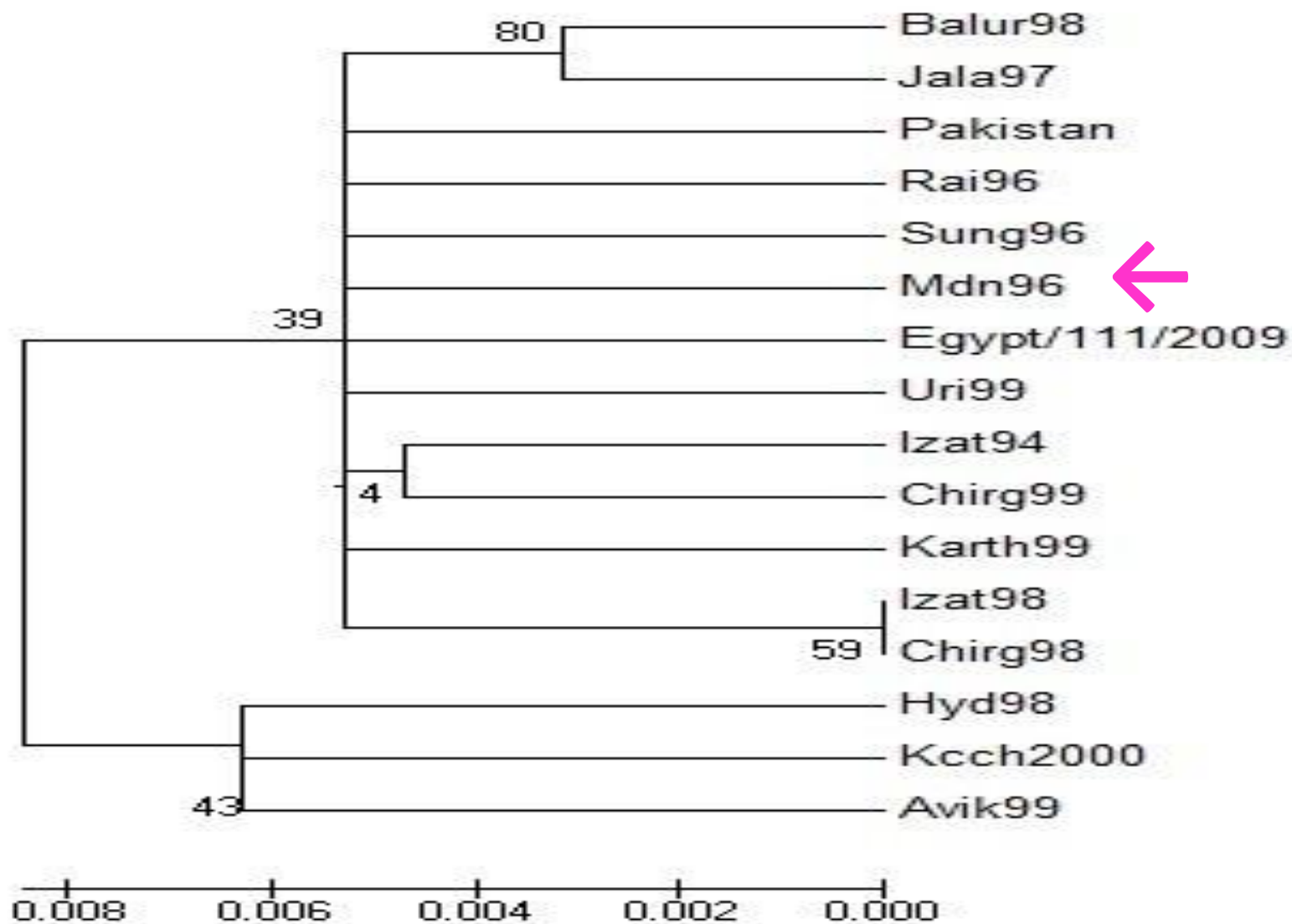
GAA	E	0.333	18.692	2+
GAG	E	0.667	37.383	4+
TTC	F	1.000	37.383	4+
TTT	F	0.000	<u>0.000</u>	0+
GGA	G	0.000	<u>0.000</u>	0+
GGC	G	1.000	9.346	1+
GGG	G	0.000	<u>0.000</u>	0+
GGT	G	0.000	<u>0.000</u>	0+
CAC	H	0.500	18.692	2+
CAT	H	0.500	18.692	2+
ATA	I	0.500	28.037	3+
ATC	I	0.333	18.692	2+
ATT	I	0.167	9.346	1+
AAA	K	0.600	28.037	3+
AAG	K	0.400	18.692	2+
CTA	L	0.000	<u>0.000</u>	0+
CTC	L	0.250	37.383	4+
CTG	L	0.312	46.729	5+
CTT	L	0.000	<u>0.000</u>	0+

TTA	L	0.000	<u>0.000</u>	0+
TTG	L	0.438	65.421	7+
ATG	M	1.000	28.037	3+
AAC	N	0.000	<u>0.000</u>	0+
AAT	N	1.000	9.346	1+
CCA	P	0.375	28.037	3+
CCC	P	0.250	18.692	2+
CCG	P	0.250	18.692	2+
CCT	P	0.125	9.346	1+
CAA	Q	0.500	37.383	4+
CAG	Q	0.500	37.383	4+
AGA	R	0.455	46.729	5+
AGG	R	0.364	37.383	4+
CGA	R	0.182	18.692	2+
CGC	R	0.000	<u>0.000</u>	0+
CGG	R	0.000	<u>0.000</u>	0+
CGT	R	0.000	<u>0.000</u>	0+
AGC	S	0.143	9.346	1+

AGT	S	0.143	9.346	1+
TCA	S	0.286	18.692	2+
TCC	S	0.286	18.692	2+
TCG	S	0.143	9.346	1+
TCT	S	0.000	<u>0.000</u>	0+
ACA	T	0.500	9.346	1+
ACC	T	0.000	<u>0.000</u>	0+
ACG	T	0.000	<u>0.000</u>	0+
ACT	T	0.500	9.346	1+
GTA	V	0.250	9.346	1+
GTC	V	0.250	9.346	1+
GTG	V	0.000	<u>0.000</u>	0+
GTT	V	0.500	18.692	2+
TGG	W	1.000	18.692	2+
TAC	Y	1.000	18.692	2+
TAT	Y	0.000	<u>0.000</u>	0+
TAA	*	0.400	37.383	4+
TAG	*	0.200	18.692	2+
TGA	*	0.400	37.383	4+

6.进化树的构建

- 对PPRV毒株进行进化分析，从NCBI核酸数据库中下载PPRV F基因的序列，共下载16条序列，并选取其中的大约为300bp大小的片段用MEGA5.05软件构建进化树，结果显示如图



7.利用WebLab统计PPRV,F蛋白20种不同氨基酸的组成

PEPSTATS of from 1 to 106

Molecular weight = 11321.06

Residues = 106

Average Residue Weight = 106.802

Charge = 5.0

Isoelectric Point = 10.4639

A280 Molar Extinction Coefficient = 1280

A280 Extinction Coefficient 1mg/ml = 0.11

Improbability of expression in inclusion bodies = 0.805

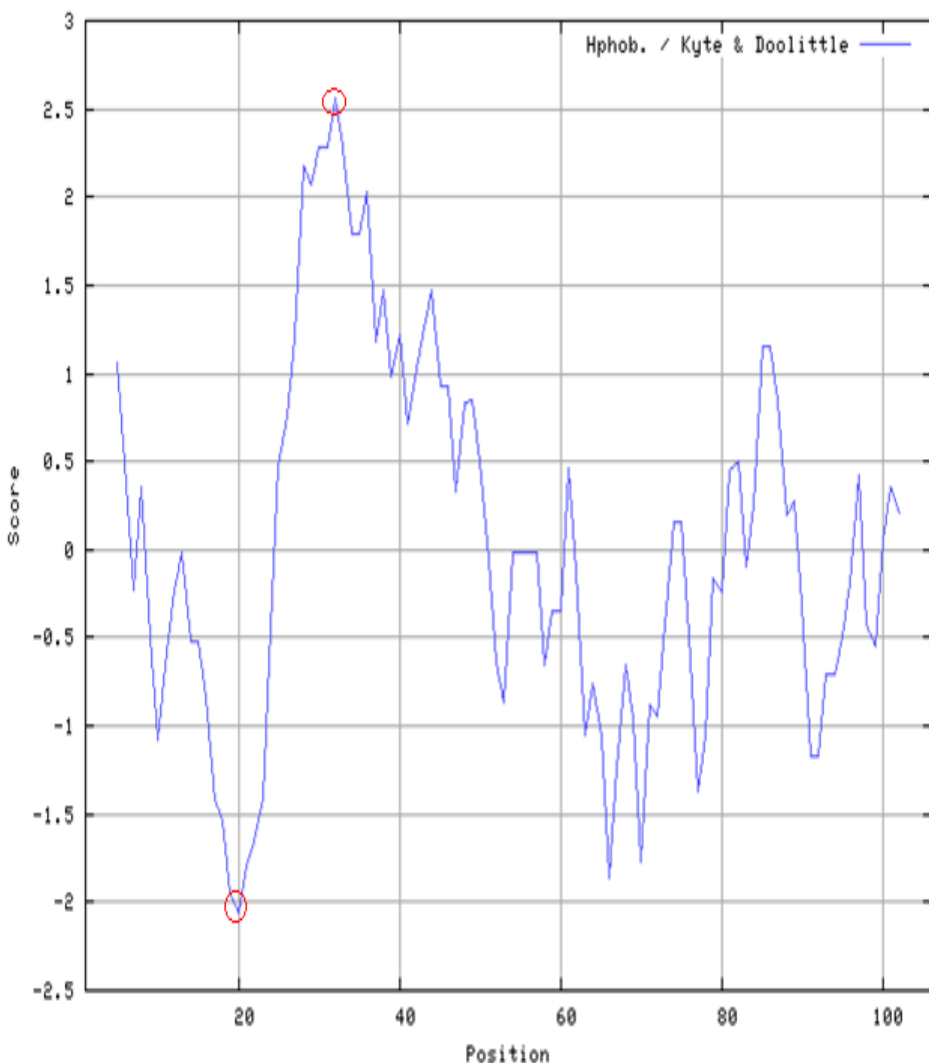
Residue	Number	Mole%	DayhoffStat
A = Ala	<u>14</u>	13.208	1.536
B = Asx	0	0.000	0.000
C = Cys	1	0.943	0.325
D = Asp	1	0.943	0.172
E = Glu	6	5.660	0.943
F = Phe	1	0.943	0.262
G = Gly	6	5.660	0.674
H = His	2	1.887	0.943
I = Ile	8	7.547	1.677
J = ---	0	0.000	0.000
K = Lys	4	3.774	0.572
L = Leu	<u>11</u>	10.377	1.402
M = Met	2	1.887	1.110
N = Asn	6	5.660	1.316
O = ---	0	0.000	0.000
P = Pro	3	2.830	0.544

Q = Gln	7	6.604	1.693
R = Arg	7	6.604	1.348
S = Ser	8	7.547	1.078
T = Thr	8	7.547	1.237
U = ---	0	0.000	0.000
V = Val	<u>10</u>	9.434	1.429
W = Trp	0	0.000	0.000
X = Xaa	0	0.000	0.000
Y = Tyr	1	0.943	0.277
Z = Glx	0	0.000	0.000

Property	Residues	Number	Mole%	
Tiny	(A+C+G+S+T)	37		34.906
Small	(A+B+C+D+G+N+P+S+T+V)	57		53.774
Aliphatic	(A+I+L+V)	43	40.566	
Aromatic	(F+H+W+Y)	4	3.774	
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	57		53.774
Polar	(D+E+H+K+N+Q+R+S+T+Z)	49		46.226
Charged	(B+D+E+H+K+R+Z)	20		18.868
Basic	(H+K+R)	13		12.264
Acidic	(B+D+E+Z)	7		6.604

8.用ProtScale预测氨基酸序列的疏水性/亲水性

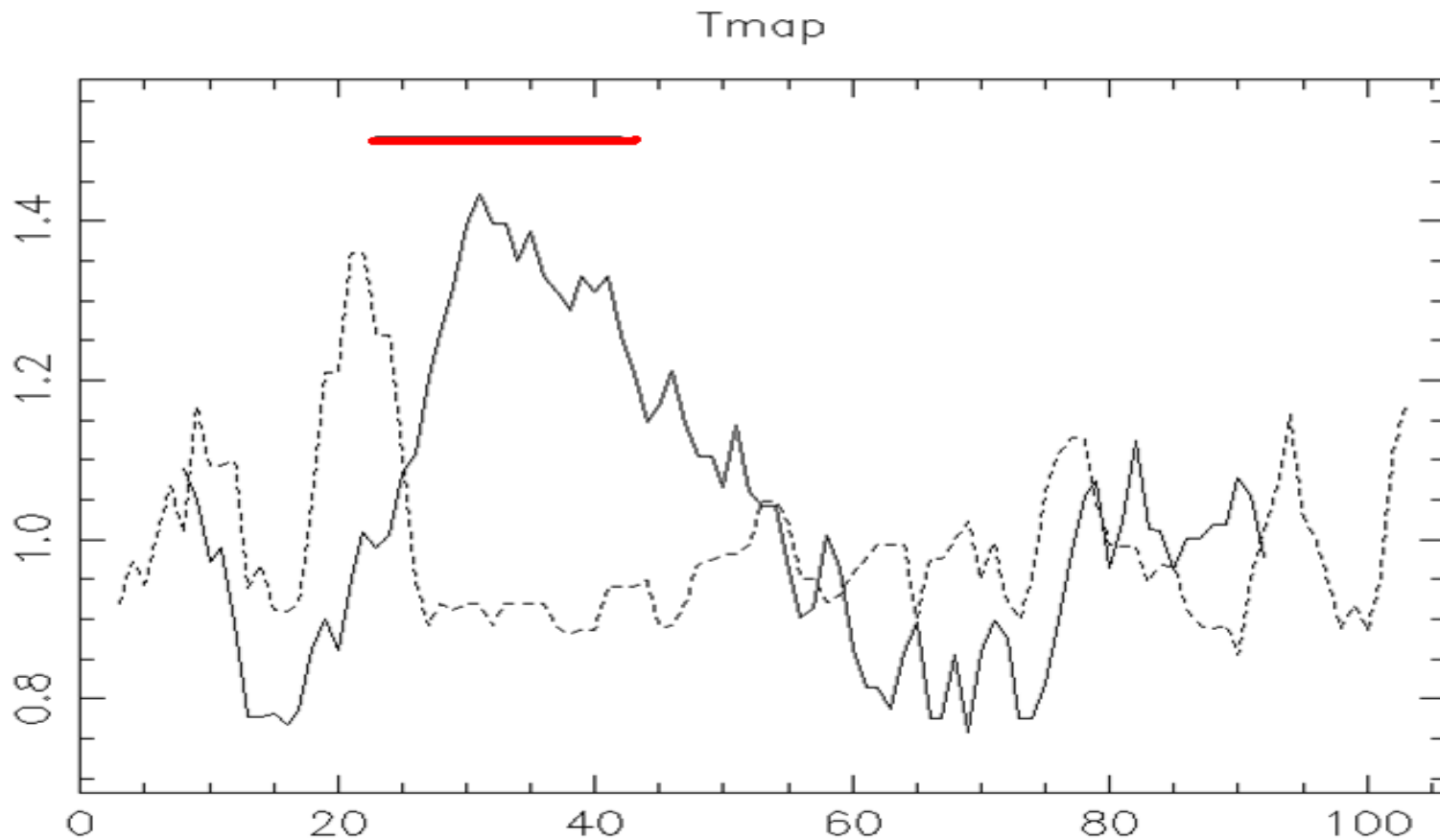
ProtScale output for user sequence



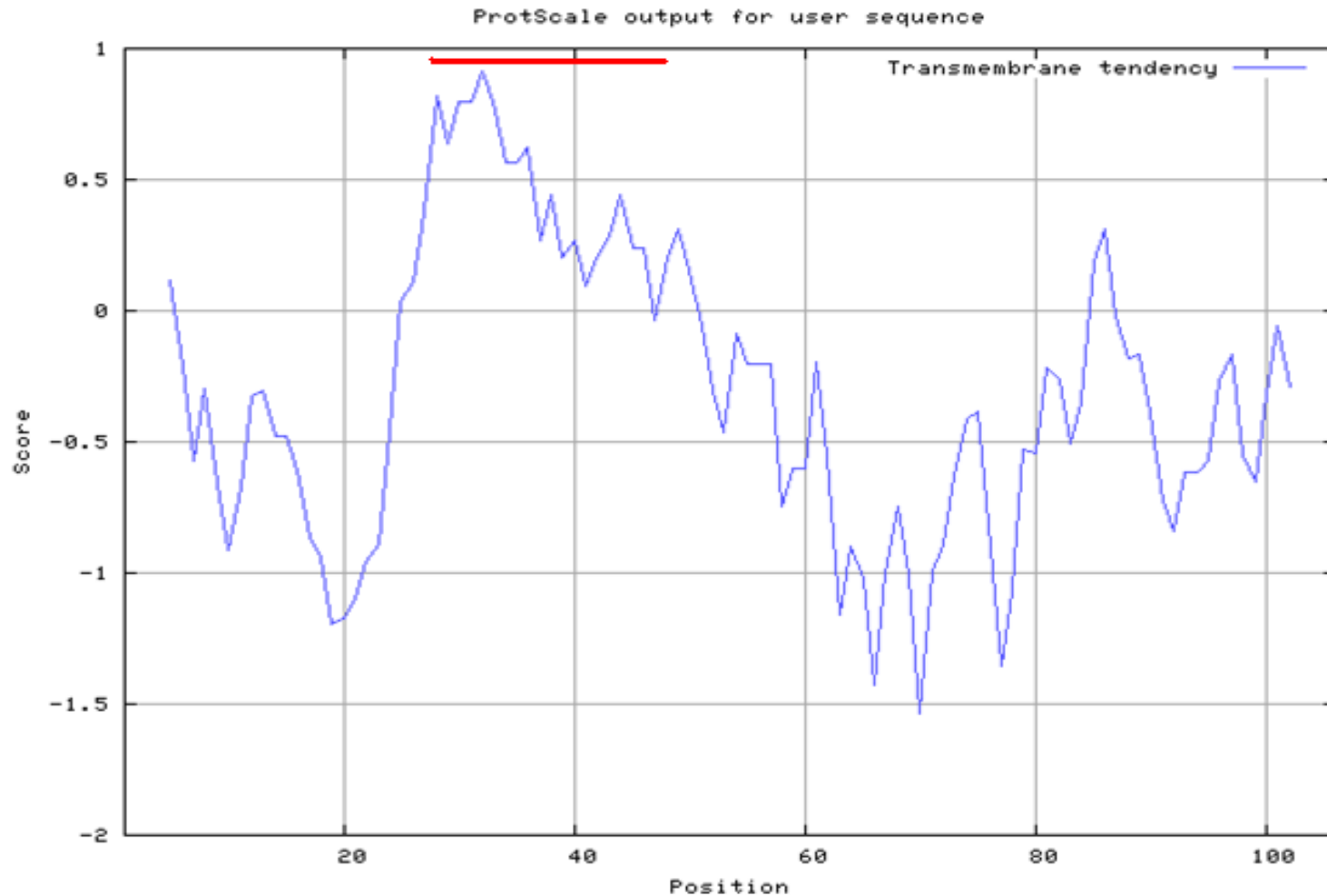
依据氨基酸分值越低亲水性越强、分值越高疏水性越强的规律，可以看出多肽链第20位具有最低的分值-1.95和最强的亲水性；第35位具有最高的分值2.58和最强的疏水性，整个多肽链表现为疏水性。

9.预测小反刍兽疫病毒F蛋白质序列中可能的跨膜螺旋

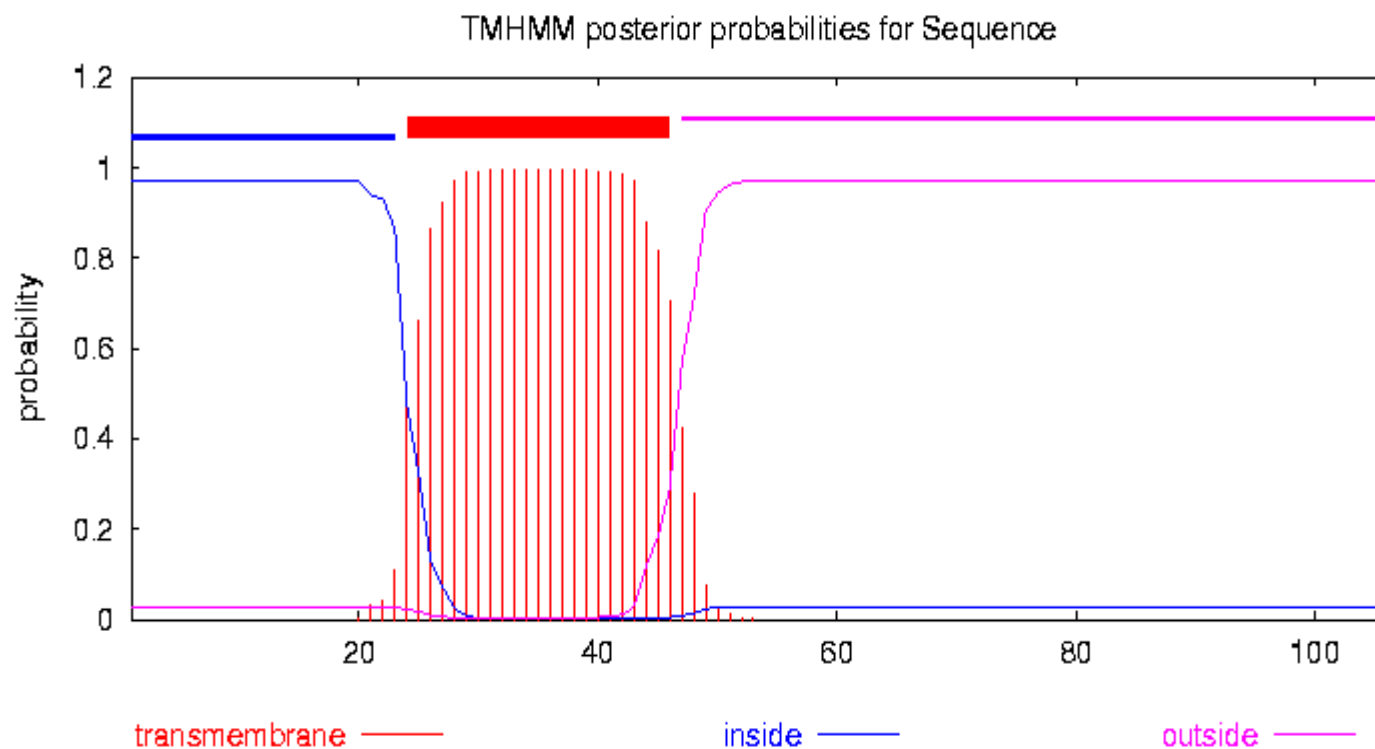
(1) 用webleb中Tmap软件程序



(2) 用 ExPASy , ProtScale , 输入氨基酸序列 , Transmembrane tendency :



(3) TMHMM

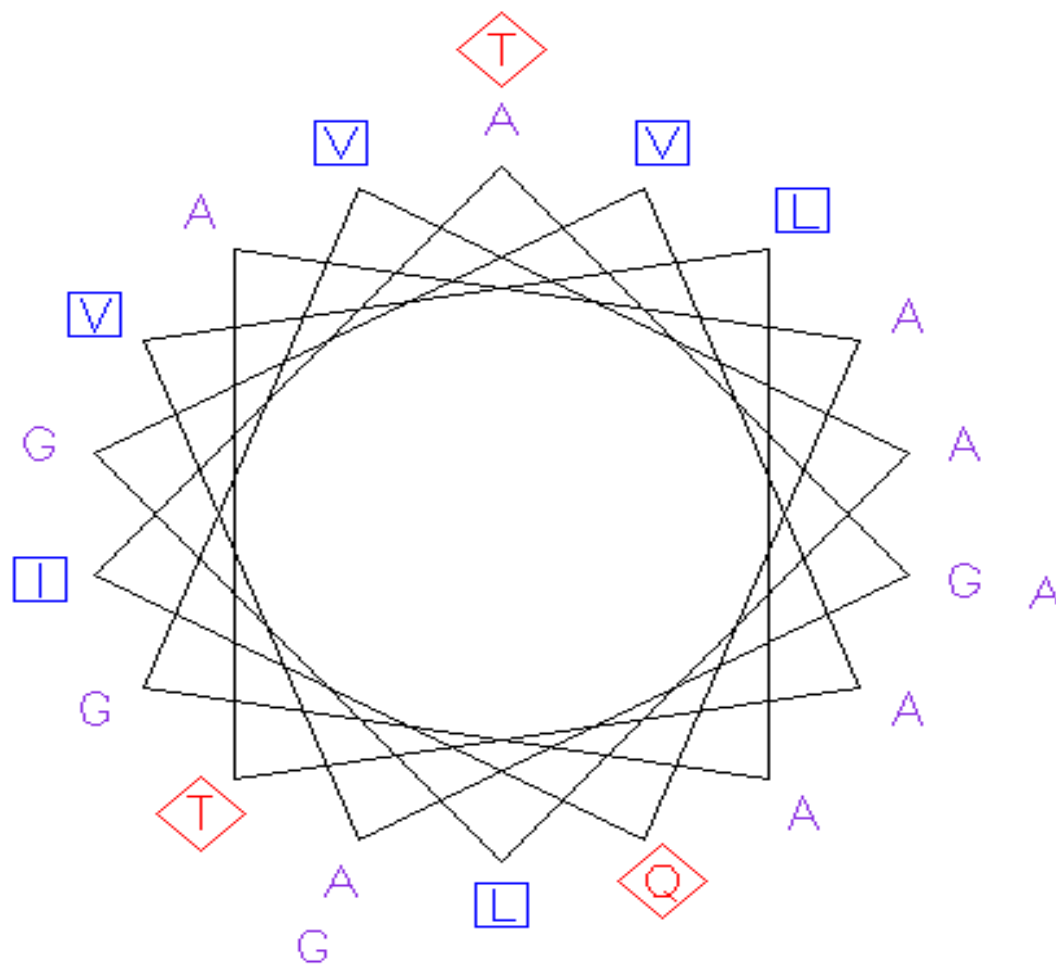


以上三个软件预测的跨膜螺旋结果基本一致，存在一个跨膜结构域。位于25-45。

用peelwheel对跨膜区绘制螺旋轮

Helical wheel of raw::543261

Thu 7 Jun 2012 10:17:52



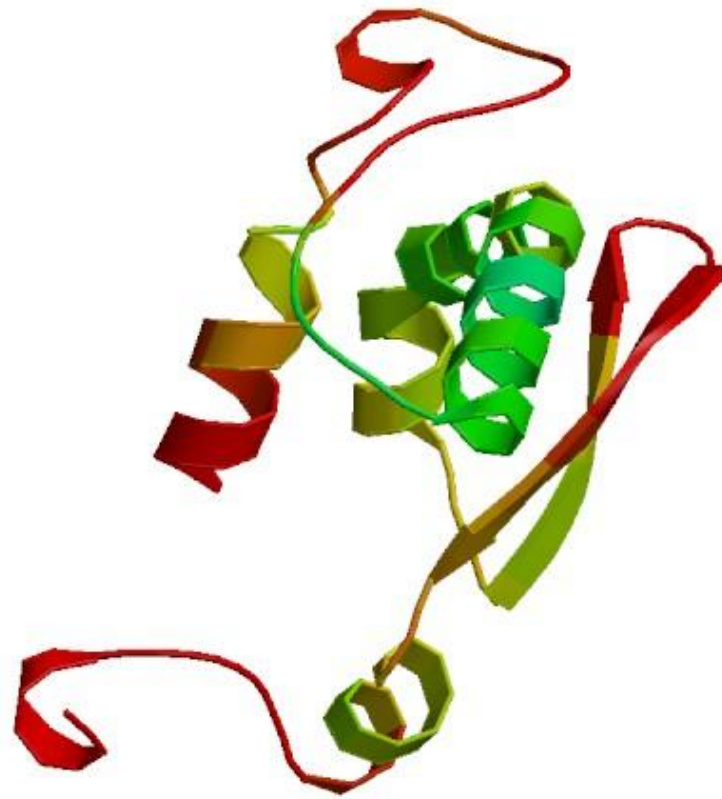
10.利用Weblab中garnier软件进行二级结构预测

```
..
#-----
      .      10      .      20      .      30      .      40      .      50
ALSVITKNVRFIQTLTPGRRTRRFAGAVLAGVALGVATAAQITAGVALHQ
helix  HH                                  HHHHHHHHHHHHHHHHHHHHHHH
sheet  EEEE      EEEEEEE      EEEEE EE
turns  TT          TTT
coil   C          CC          C
      .      60      .      70      .      80      .      90      .      100
SLMNSQAIESLKTSLEKSNQAIEEIRLANKETILAVQGVDYINNELVFS
helix  HHHHHHHHHHHHHHH HHHHHHHHHHHHHHHHH H
sheet  E EEEEEEEEEEE E
turns  T          T
coil   CCCC      C          C CC
      O
      VHRMSC
helix  HH
sheet
turns  TTTT
coil
#-----
..
```

可以看出在30—50和55—67位以及70—84位AA处有螺旋区域

蛋白质的三级结构预测

- 用Swiss-model将*F Protein in PPRV*与蛋白质结构数据库中的蛋白质三维结构进行匹配，以PARAINFLUENZA VIRUS 5 F PROTEIN (79-185)为模版，输出模拟的三维结构图。



Estimated per-residue inaccuracy visualised using a colour gradient from **green(more reliable regions) to **red**(potentially unreliable regions)**





— PARAINFLUENZA
VIRUS

— PPRV

Reference

- Benkert P, Biasini M, Schwede T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics*, 27(3):343–50.
- Arnold K., Bordoli L., Kopp J., Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195–201.
- Schwede T, Kopp J, Guex N, and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31: 3381–3385.

- ◆ **首先，感谢罗老师这个学期对我们学习的指导！**
- ◆ **其次，感谢我的组员刘静宜、刘伟、李君君三位同学的努力、协作！**
- ◆ **最后，感谢研究生院给我们提供的学习平台！**

谢谢观赏