猪瘟病毒E2蛋白序列、结构及其功能分析

Seqeunce, structure and function analysis of CSFV E2

小组：S1G05

组员：岳龙 王素艳 蔺俊丽 李婷婷

报告人：李婷婷

主要内容

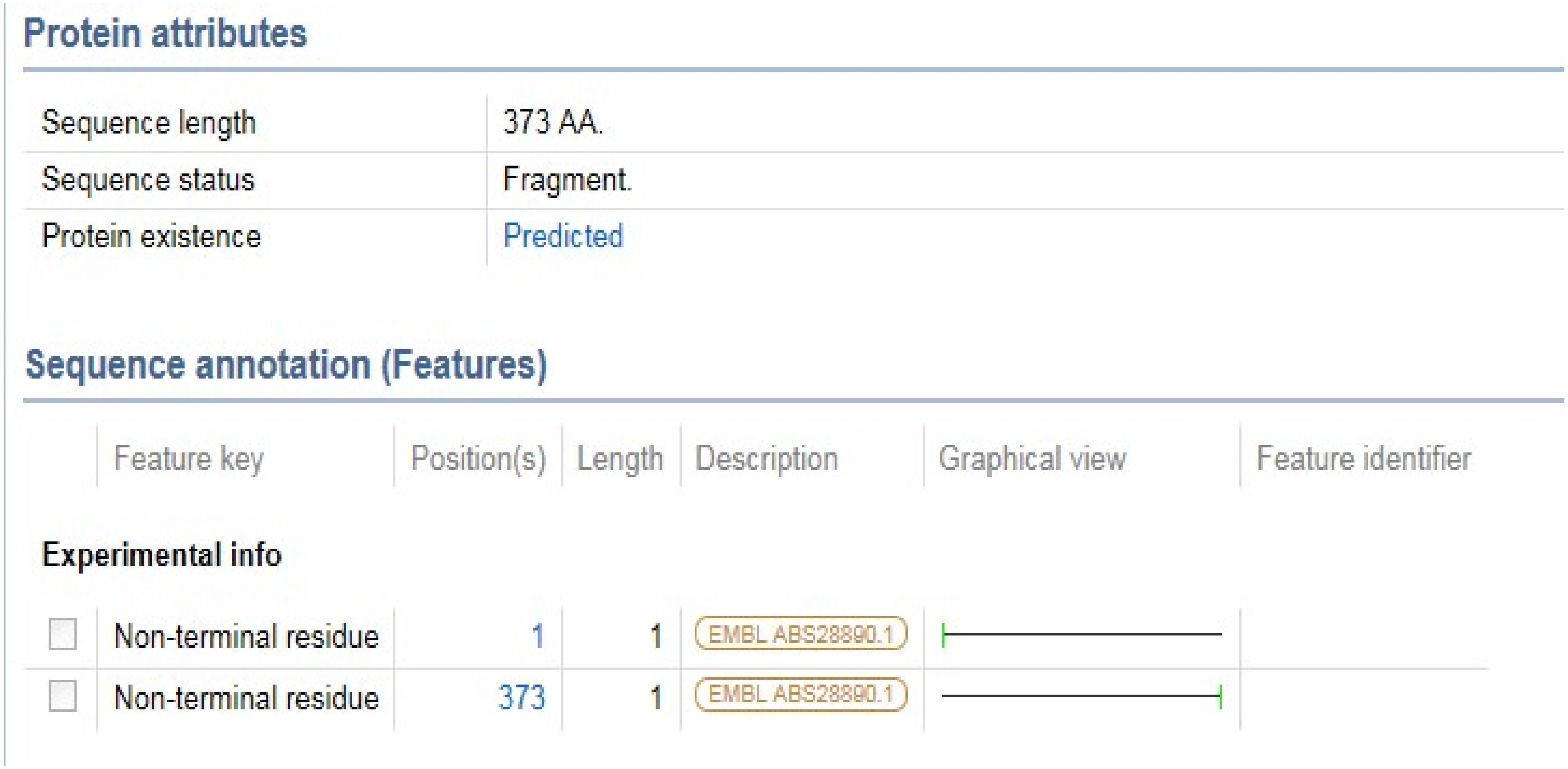
- 一、研究背景
- 二、E2蛋白的核酸序列分析
- 三、E2蛋白的氨基酸序列分析
- 四、E2蛋白的结构分析

研究背景

- 猪瘟病毒（CSFV）是黄病毒科瘟病毒属成员，瘟病毒属还包括牛病毒性腹泻病毒（BVDV）和绵羊边界病毒（BDV）。基因组为单股正链RNA，包含一个大的开放阅读框，编码四个结构蛋白（C、E0、E1、E2）和8个非结构蛋白。
- E2蛋白是病毒诱导机体产生中和抗体的主要保护性抗原，因此，研究清楚它的序列和结构对研制高效疫苗和快速血清学诊断试剂具有重大意义。

E2蛋白核酸序列分析

从uniprot调取E2蛋白的氨基酸序列，登录号A7L897。



Protein attributes

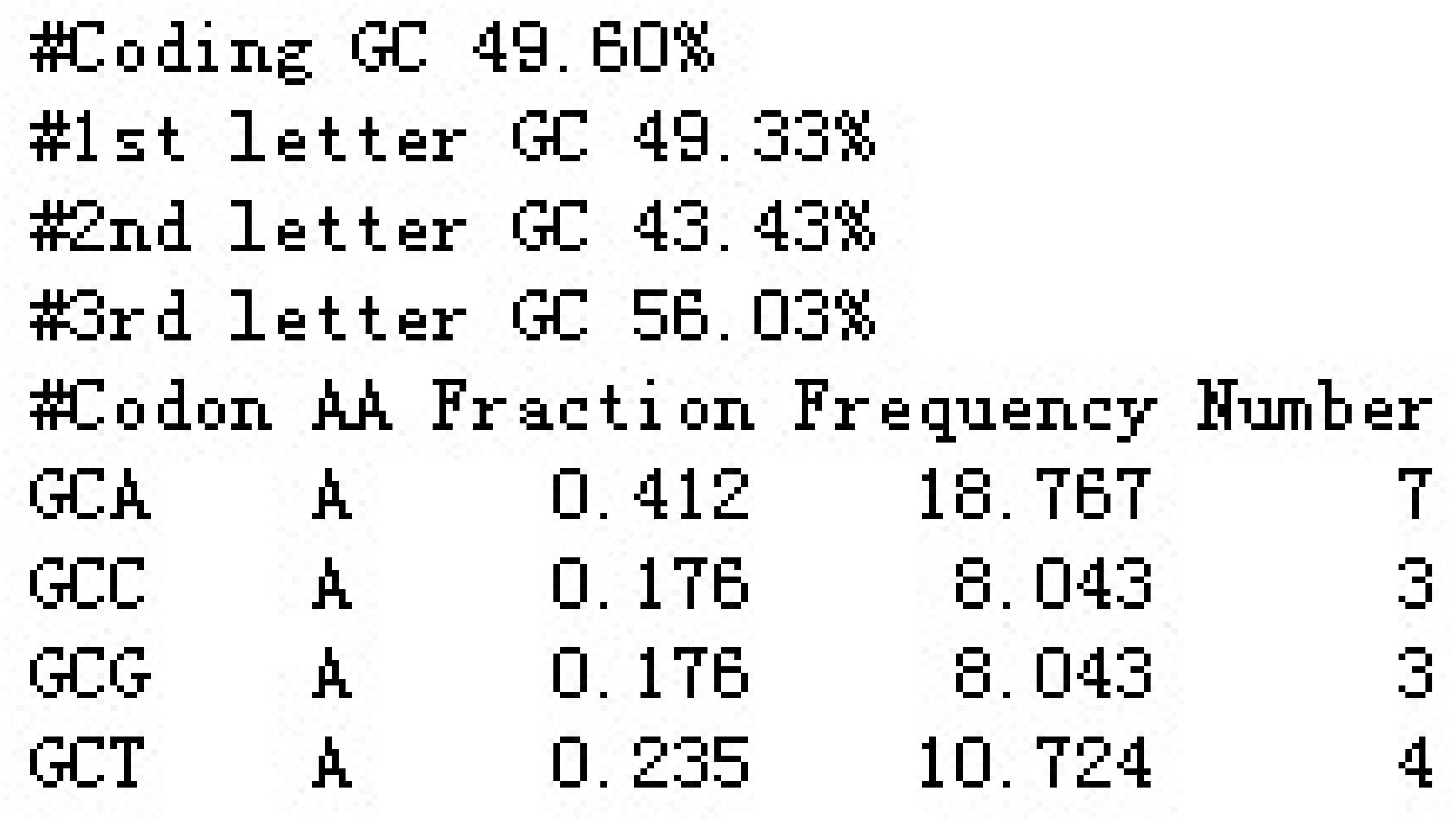
Sequence length	373 AA.
Sequence status	Fragment.
Protein existence	Predicted

Sequence annotation (Features)

	Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Experimental info						
☐	Non-terminal residue	1	1	EMBL ABS28890.1		
☐	Non-terminal residue	373	1	EMBL ABS28890.1		

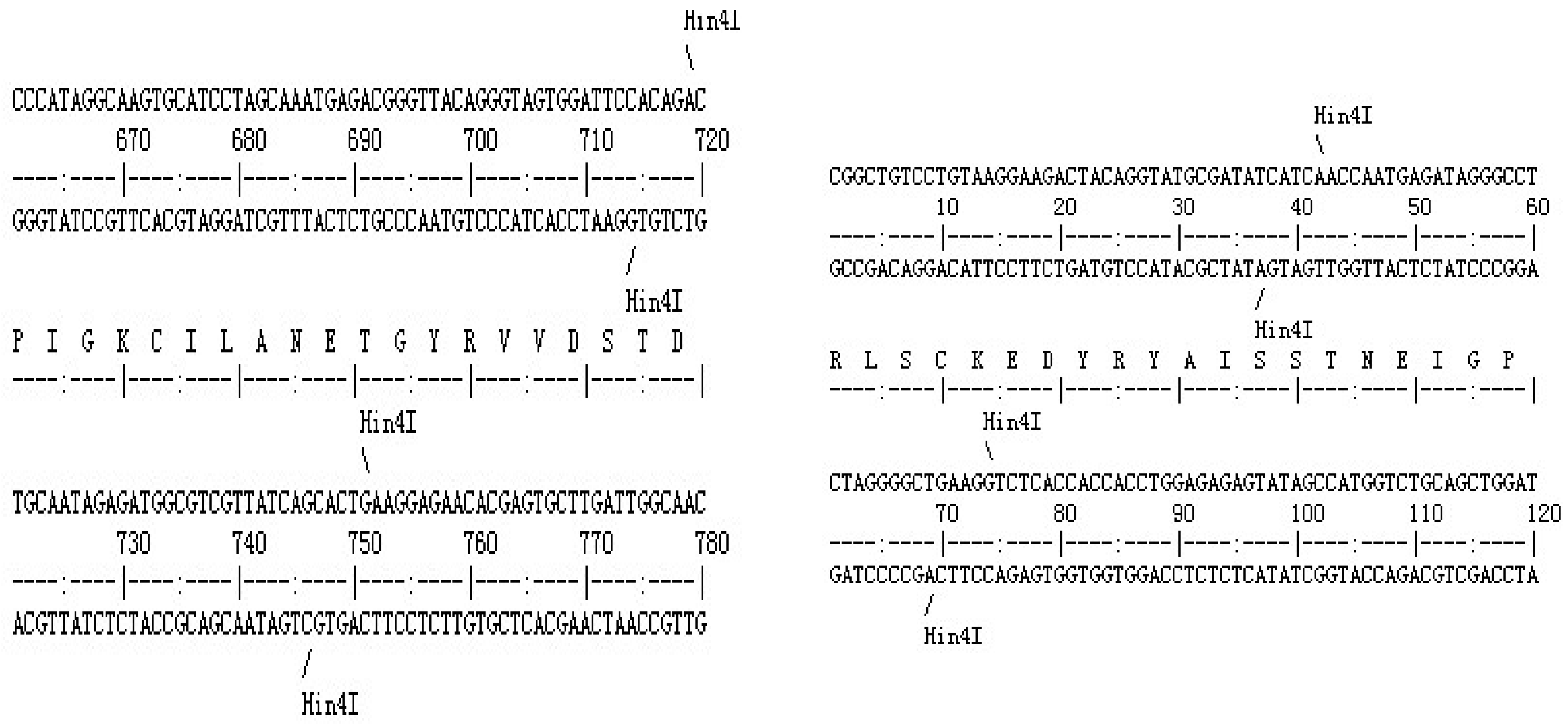
E2蛋白核苷酸序列分析

利用cusp程序分析该基因的密码子使用特征



```
#Coding GC 49.60%
#1st letter GC 49.33%
#2nd letter GC 43.43%
#3rd letter GC 56.03%
#Codon AA Fraction Frequency Number
GCA    A      0.412    18.767      7
GCC    A      0.176     8.043      3
GCG    A      0.176     8.043      3
GCT    A      0.235    10.724      4
```

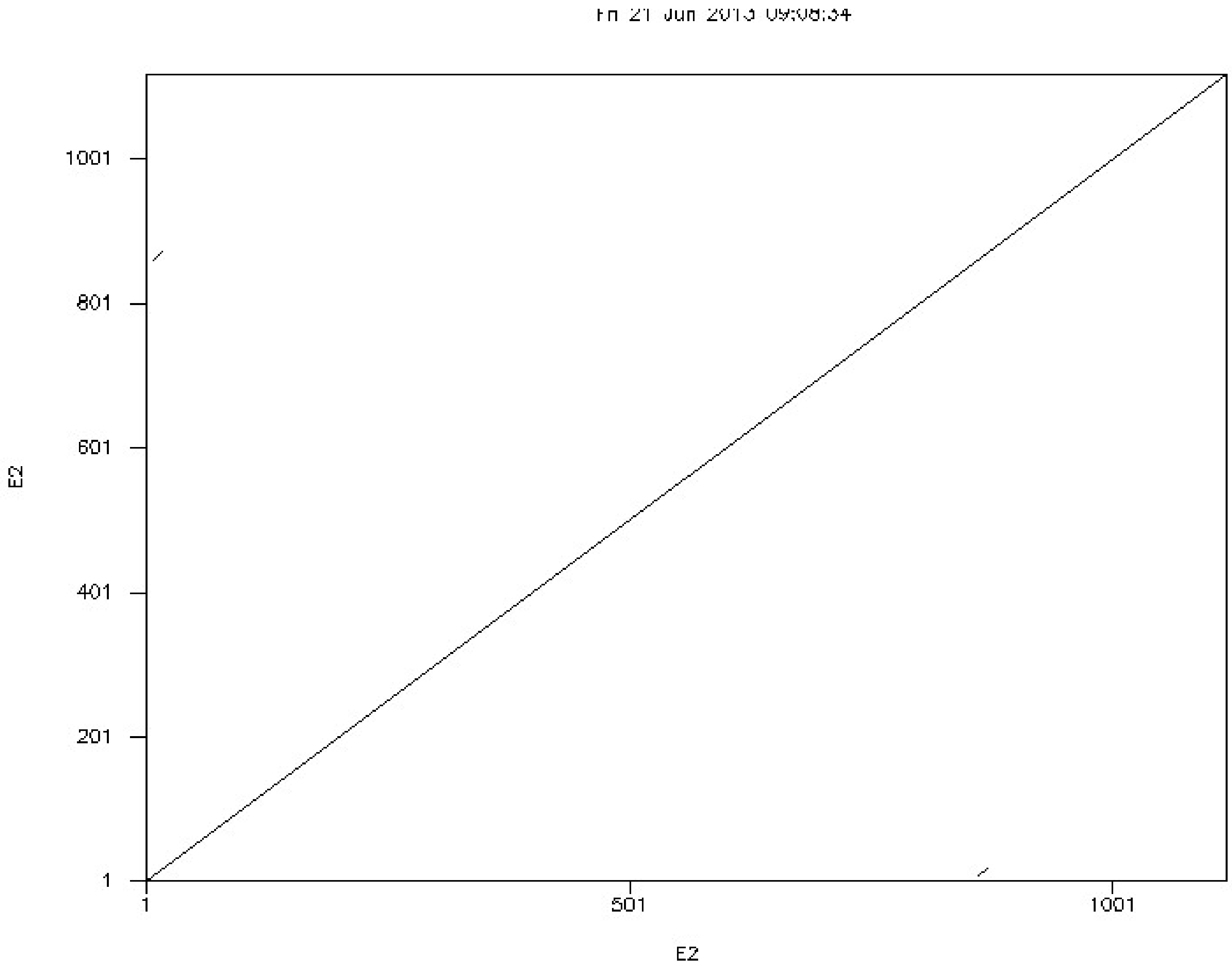
利用remap程序分析酶切位点，有限制性内切酶Hin4I的2个酶切位点



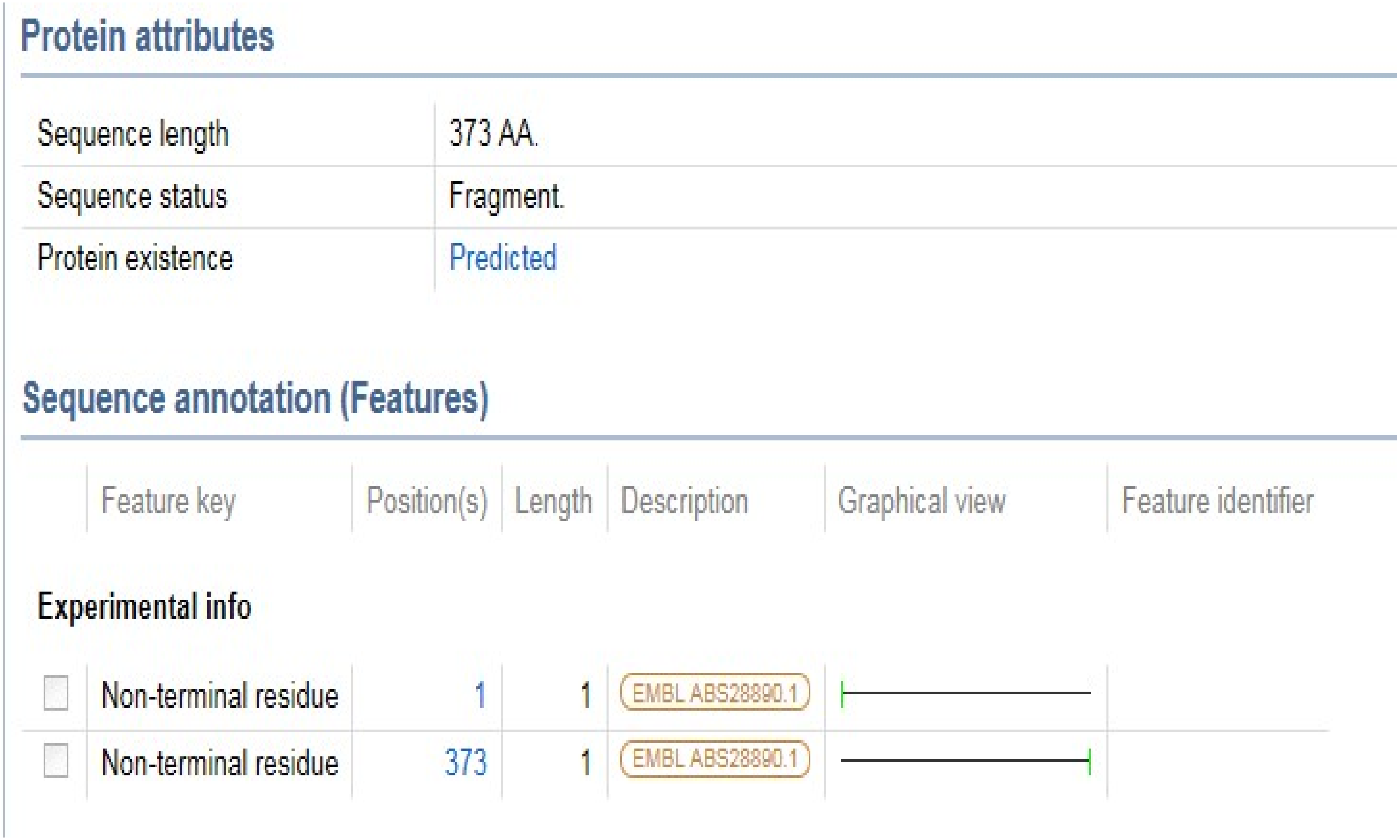
```
                                                          Hin4I
                                                          \
CCCATAGGCAAGTGCATCCTAGCAAATGAGACGGGTTACAGGGTAGTGGATTCCACAGAC
        670       680       690       700       710       720
----:----|----:----|----:----|----:----|----:----|----:----|
GGGTATCCGTTCACGTAGGATCGTTTACTCTGCCCAATGTCCCATCACCTAAGGTGTCTG
                                                     /
                                                     Hin4I

P  I  G  K  C  I  L  A  N  E  T  G  Y  R  V  V  D  S  T  D
----:----|----:----|----:----|----:----|----:----|----:----|
                              Hin4I
                              \
TGCAATAGAGATGGCGTCGTTATCAGCACTGAAGGAGAACACGAGTGCTTGATTGGCAAC
        730       740       750       760       770       780
----:----|----:----|----:----|----:----|----:----|----:----|
ACGTTATCTCTACCGCAGCAATAGTCGTGACTTCCTCTTGTGCTCACGAACTAACCGTTG
                         /
                         Hin4I
```

```
                                         Hin4I
                                         \
CGGCTGTCCTGTAAGGAAGACTACAGGTATGCGATATCATCAACCAATGAGATAGGGCCT
         10        20        30        40        50        60
----:----|----:----|----:----|----:----|----:----|----:----|
GCCGACAGGACATTCCTTCTGATGTCCATACGCTATAGTAGTTGGTTACTCTATCCCGGA
                                    /
                                    Hin4I

R  L  S  C  K  E  D  Y  R  Y  A  I  S  S  T  N  E  I  G  P
----:----|----:----|----:----|----:----|----:----|----:----|
             Hin4I
             \
CTAGGGGCTGAAGGTCTCACCACCACCTGGAGAGAGTATAGCCATGGTCTGCAGCTGGAT
         70        80        90       100       110       120
----:----|----:----|----:----|----:----|----:----|----:----|
GATCCCCGACTTCCAGAGTGGTGGTGGACCTCTCTCATATCGGTACCAGACGTCGACCTA
        /
        Hin4I
```

Dottup分析



E2蛋白氨基酸序列分析



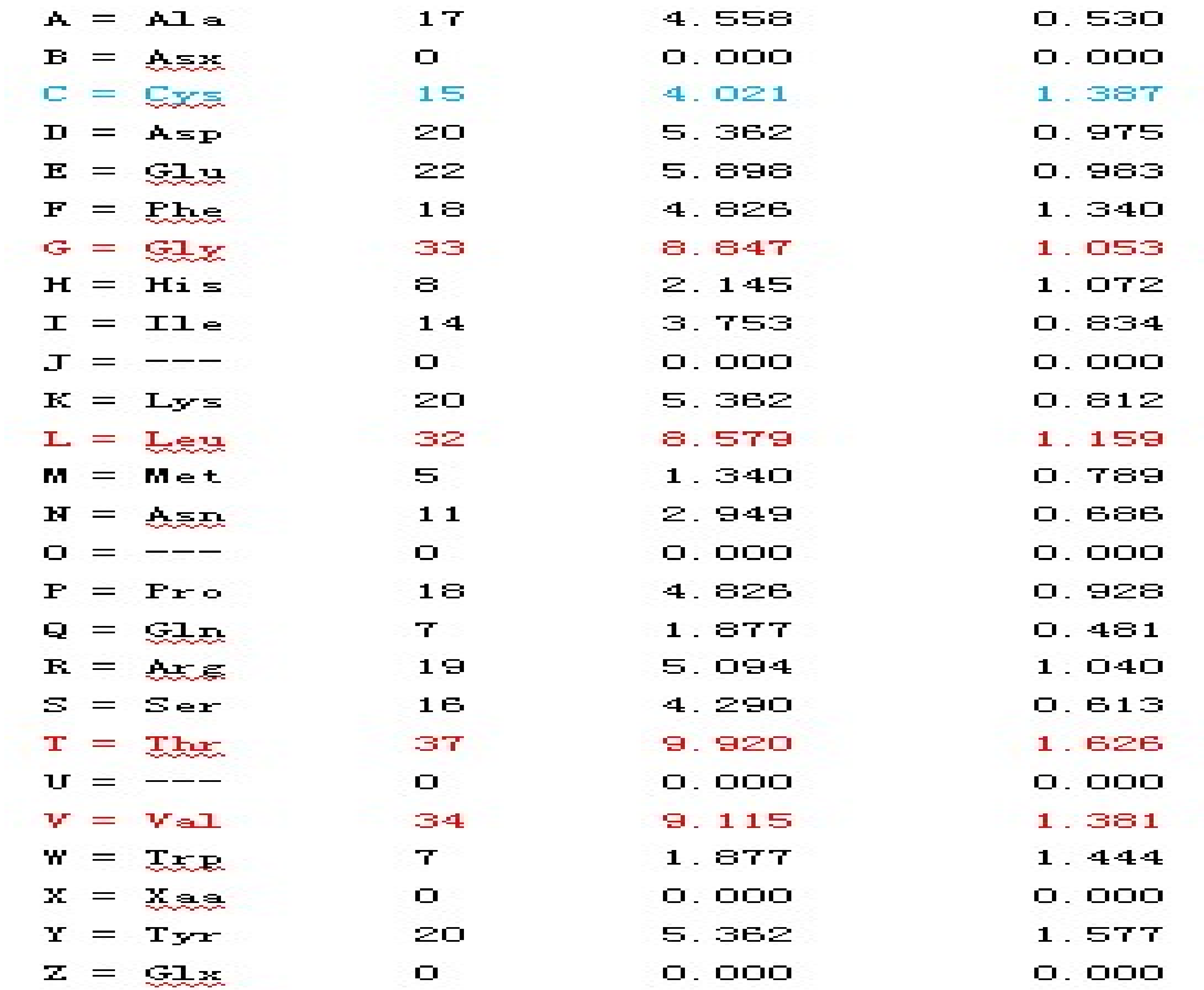
Protein attributes

Sequence length	373 AA.
Sequence status	Fragment.
Protein existence	Predicted

Sequence annotation (Features)

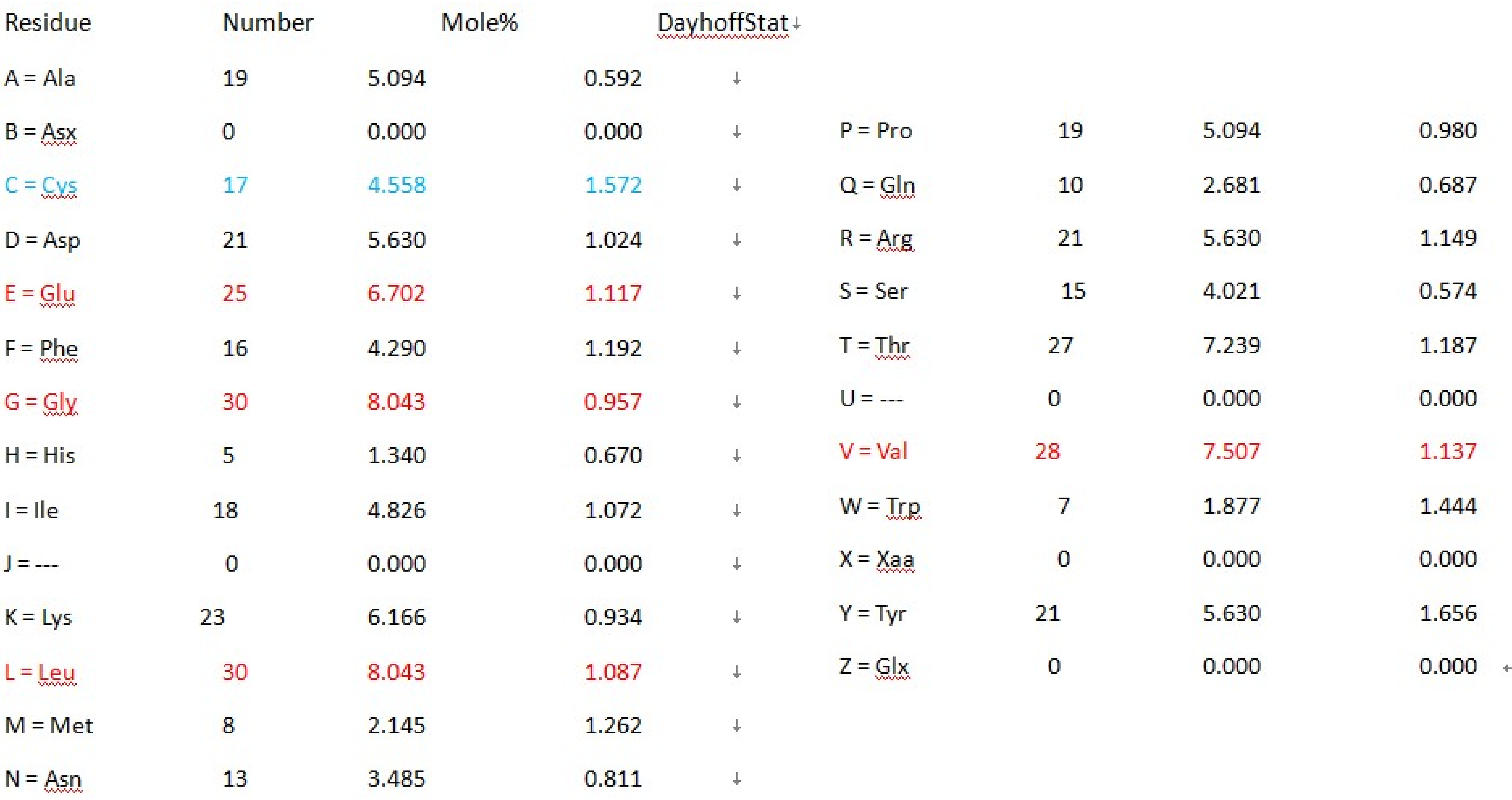
	Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Experimental info						
☐	Non-terminal residue	1	1	EMBL ABS28890.1		
☐	Non-terminal residue	373	1	EMBL ABS28890.1		

用Weblab中的氨基酸分析程序pepstats，分析此蛋白的氨基酸组成



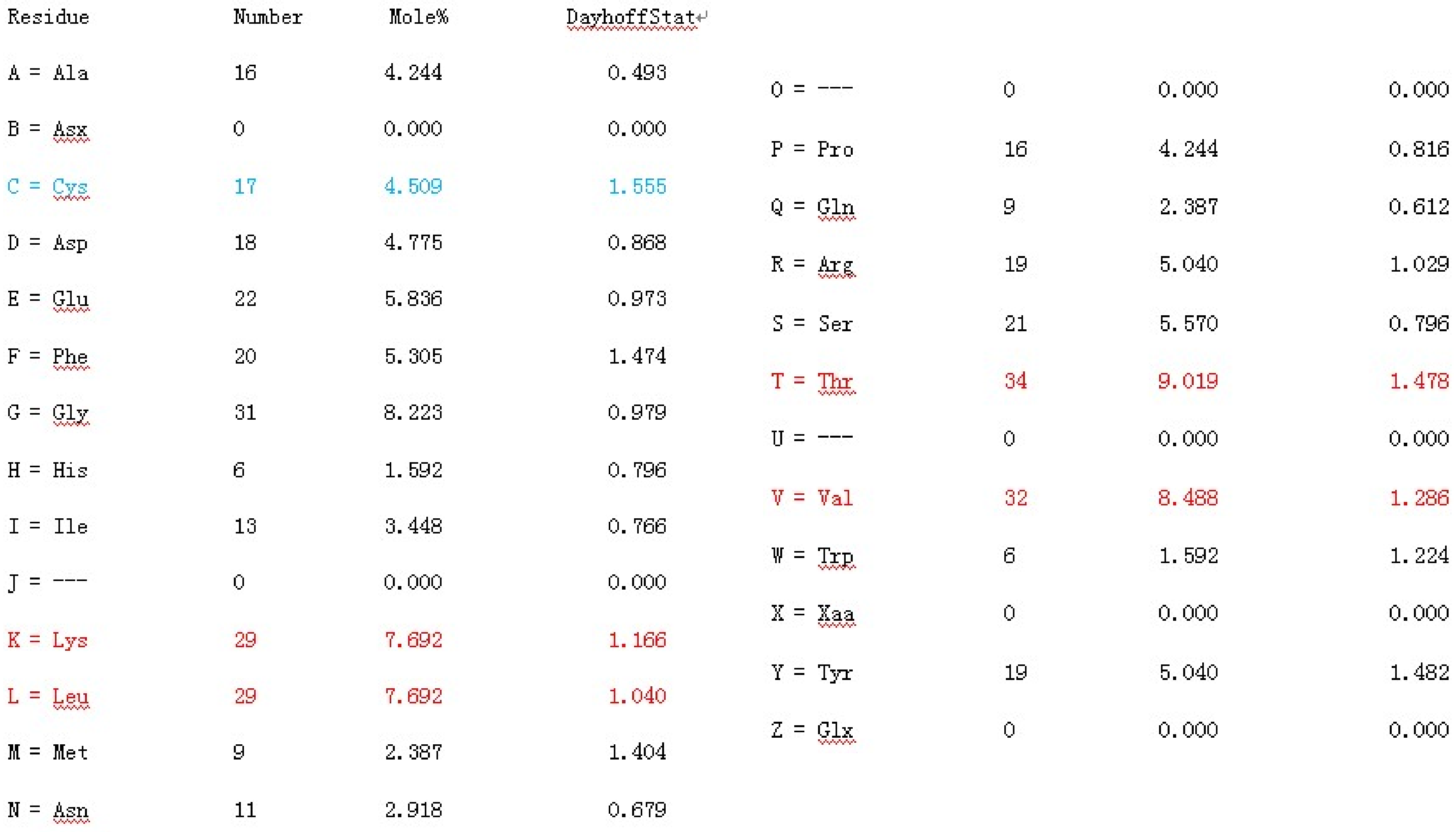
```
A = Ala        17        4.558        0.530
B = Asx        0         0.000        0.000
C = Cys        15        4.021        1.387
D = Asp        20        5.362        0.975
E = Glu        22        5.898        0.983
F = Phe        18        4.826        1.340
G = Gly        33        8.847        1.053
H = His        8         2.145        1.072
I = Ile        14        3.753        0.834
J = ---        0         0.000        0.000
K = Lys        20        5.362        0.812
L = Leu        32        8.579        1.159
M = Met        5         1.340        0.789
N = Asn        11        2.949        0.686
O = ---        0         0.000        0.000
P = Pro        18        4.826        0.928
Q = Gln        7         1.877        0.481
R = Arg        19        5.094        1.040
S = Ser        16        4.290        0.613
T = Thr        37        9.920        1.626
U = ---        0         0.000        0.000
V = Val        34        9.115        1.381
W = Trp        7         1.877        1.444
X = Xaa        0         0.000        0.000
Y = Tyr        20        5.362        1.577
Z = Glx        0         0.000        0.000
```

BVDVE2蛋白的氨基酸组成



Residue	Number	Mole%	DayhoffStat
A = Ala	19	5.094	0.592
B = Asx	0	0.000	0.000
C = Cys	17	4.558	1.572
D = Asp	21	5.630	1.024
E = Glu	25	6.702	1.117
F = Phe	16	4.290	1.192
G = Gly	30	8.043	0.957
H = His	5	1.340	0.670
I = Ile	18	4.826	1.072
J = ---	0	0.000	0.000
K = Lys	23	6.166	0.934
L = Leu	30	8.043	1.087
M = Met	8	2.145	1.262
N = Asn	13	3.485	0.811
P = Pro	19	5.094	0.980
Q = Gln	10	2.681	0.687
R = Arg	21	5.630	1.149
S = Ser	15	4.021	0.574
T = Thr	27	7.239	1.187
U = ---	0	0.000	0.000
V = Val	28	7.507	1.137
W = Trp	7	1.877	1.444
X = Xaa	0	0.000	0.000
Y = Tyr	21	5.630	1.656
Z = Glx	0	0.000	0.000

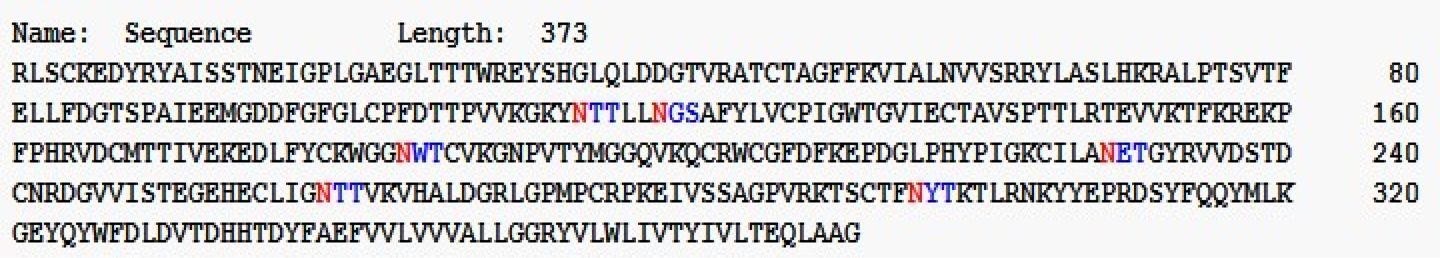
BDVE2蛋白氨基酸组成



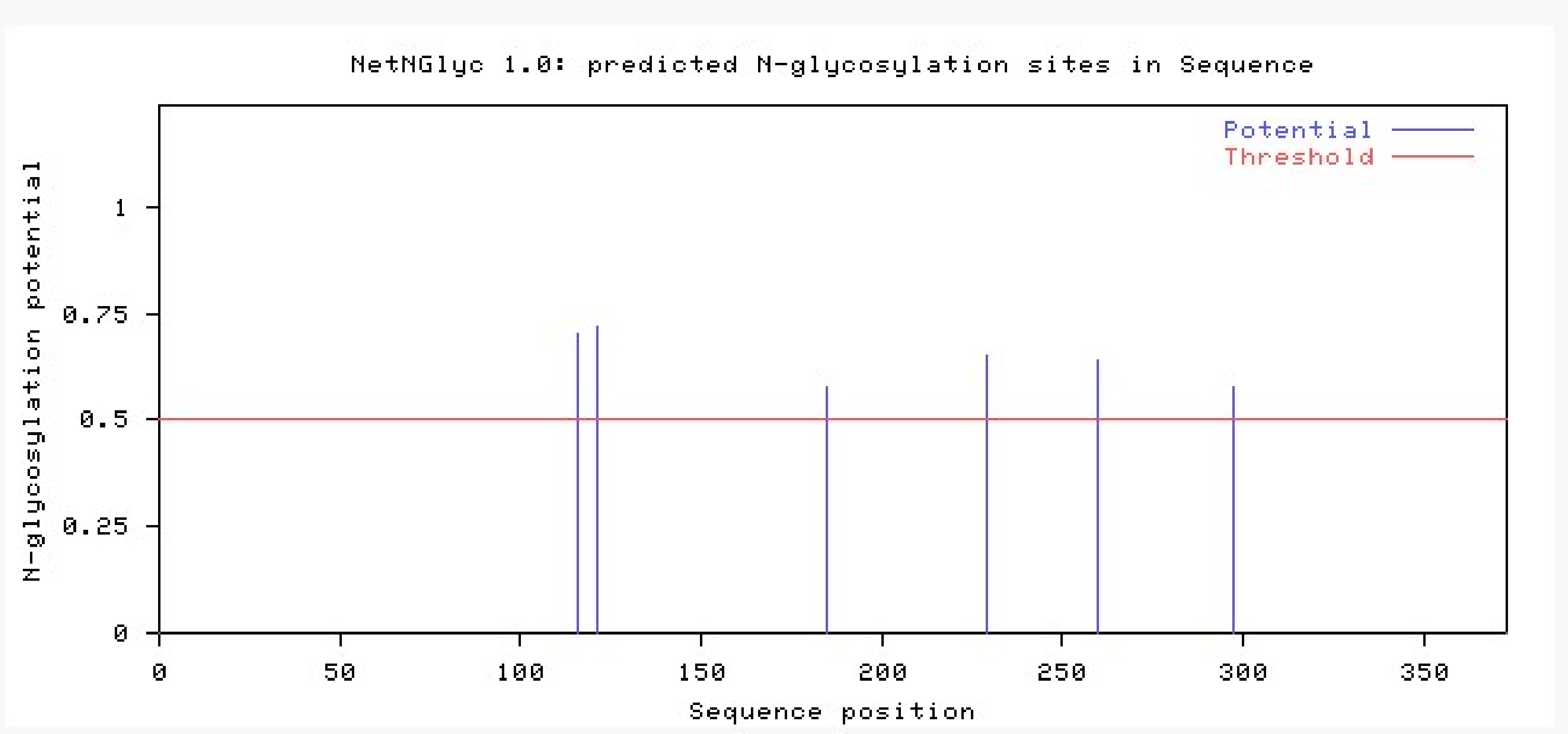
Residue	Number	Mole%	DayhoffStat
A = Ala	16	4.244	0.493
B = Asx	0	0.000	0.000
C = Cys	17	4.509	1.555
D = Asp	18	4.775	0.868
E = Glu	22	5.836	0.973
F = Phe	20	5.305	1.474
G = Gly	31	8.223	0.979
H = His	6	1.592	0.796
I = Ile	13	3.448	0.766
J = ---	0	0.000	0.000
K = Lys	29	7.692	1.166
L = Leu	29	7.692	1.040
M = Met	9	2.387	1.404
N = Asn	11	2.918	0.679
O = ---	0	0.000	0.000
P = Pro	16	4.244	0.816
Q = Gln	9	2.387	0.612
R = Arg	19	5.040	1.029
S = Ser	21	5.570	0.796
T = Thr	34	9.019	1.478
U = ---	0	0.000	0.000
V = Val	32	8.488	1.286
W = Trp	6	1.592	1.224
X = Xaa	0	0.000	0.000
Y = Tyr	19	5.040	1.482
Z = Glx	0	0.000	0.000

利用CBS网站中的糖基化位点预测程序

NetNGlyc 1.0 Server

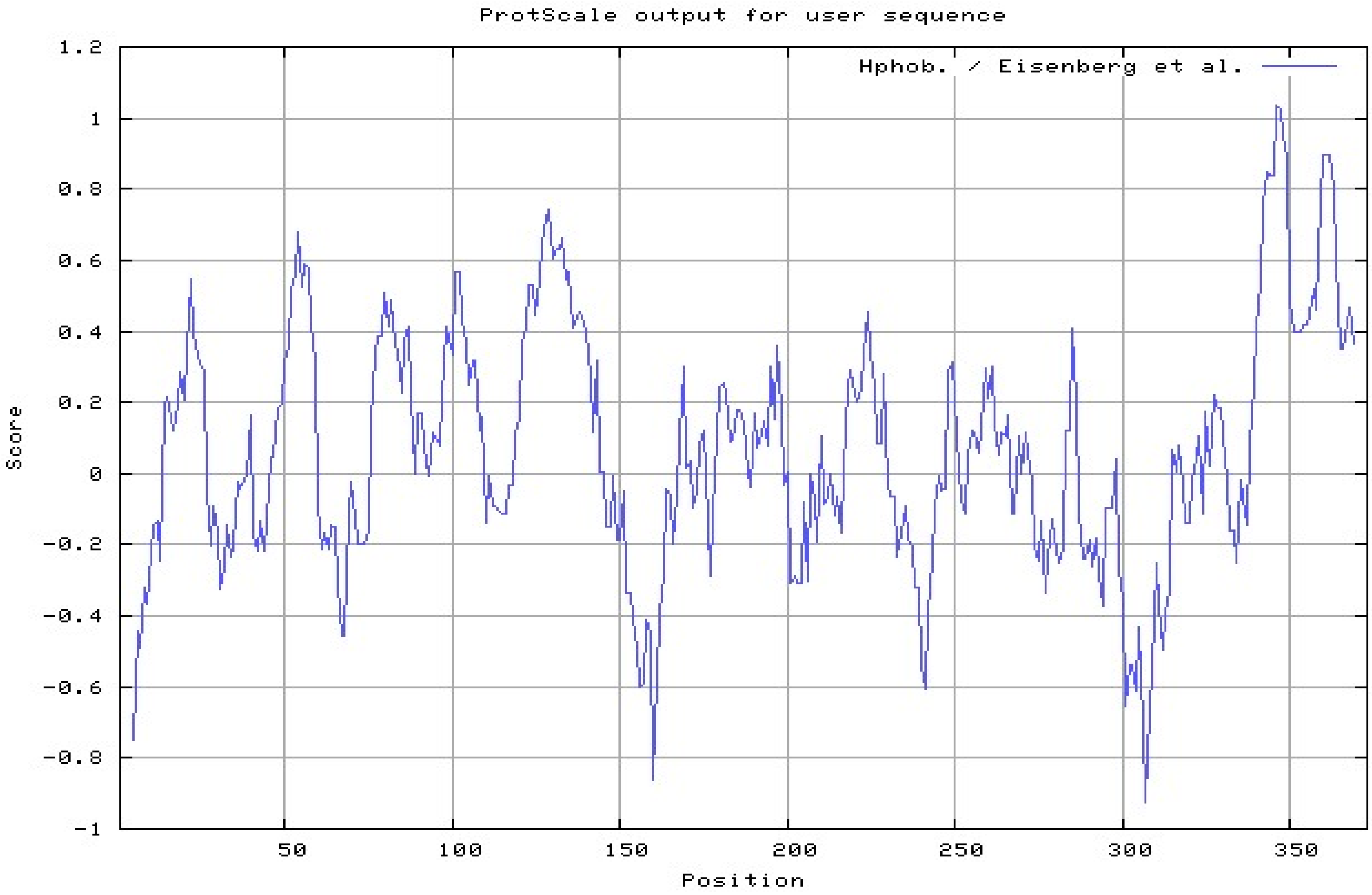


```
Name:  Sequence          Length:  373
RLSCKEDYRYAISSTNEIGPLGAEGLTTTWREYSHGLQLDDGTVRATCTAGFFKVIALNVVSRRYLASLHKRALPTSVTF      80
ELLFDGTSPAIEEMGDDFGFGLCPFDTTPVVKGKYNTTLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFKREKP     160
FPHRVDCMTTIVEKEDLFYCKWGGNWTCVKGNPVTYMGGQVKQCRWCGFDFKEPDGLPHYPIGKCILANETGYRVVDSTD     240
CNRDGVVISTEGEHECLIGNTTVKVHALDGRLGPMPCRPKEIVSSAGPVRKTSCTFNYTKTLRNKYYEPRDSYFQQYMLK     320
GEYQYWFDLDVTDHHTDYFAEFVVLVVVALLGGRYVLWLIVTYIVLTEQLAAG
```

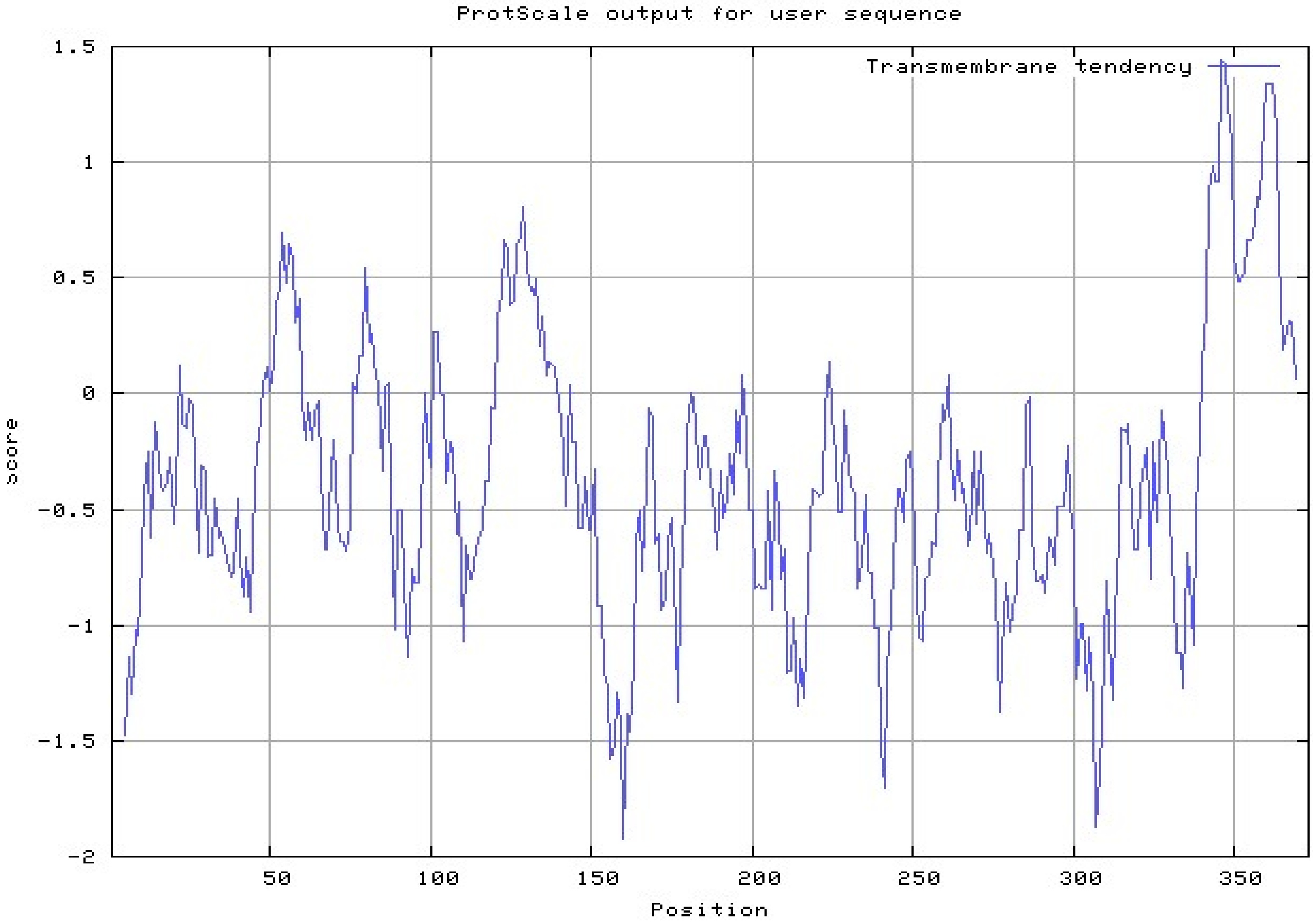


E2蛋白结构分析

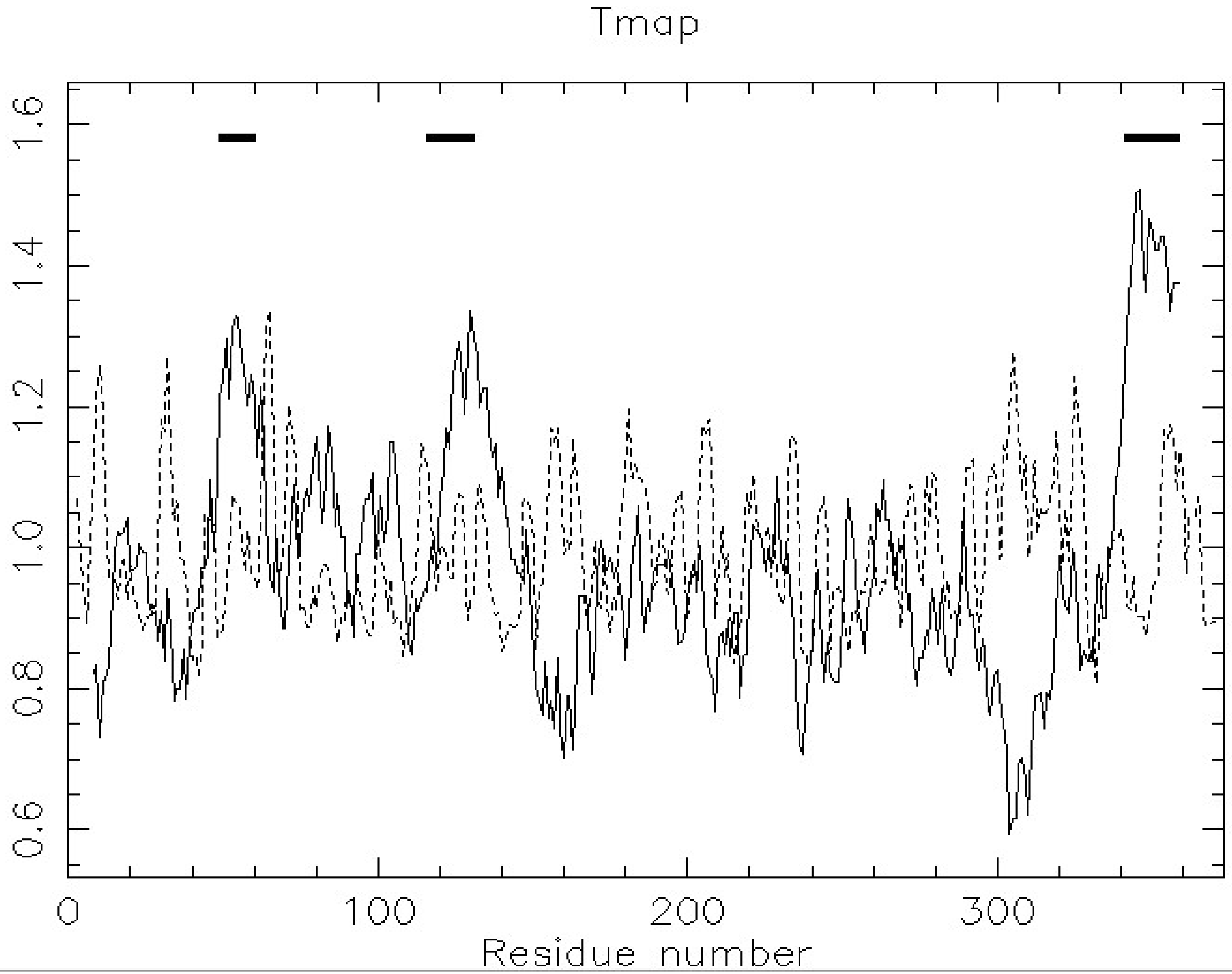
蛋白亲疏水性分析



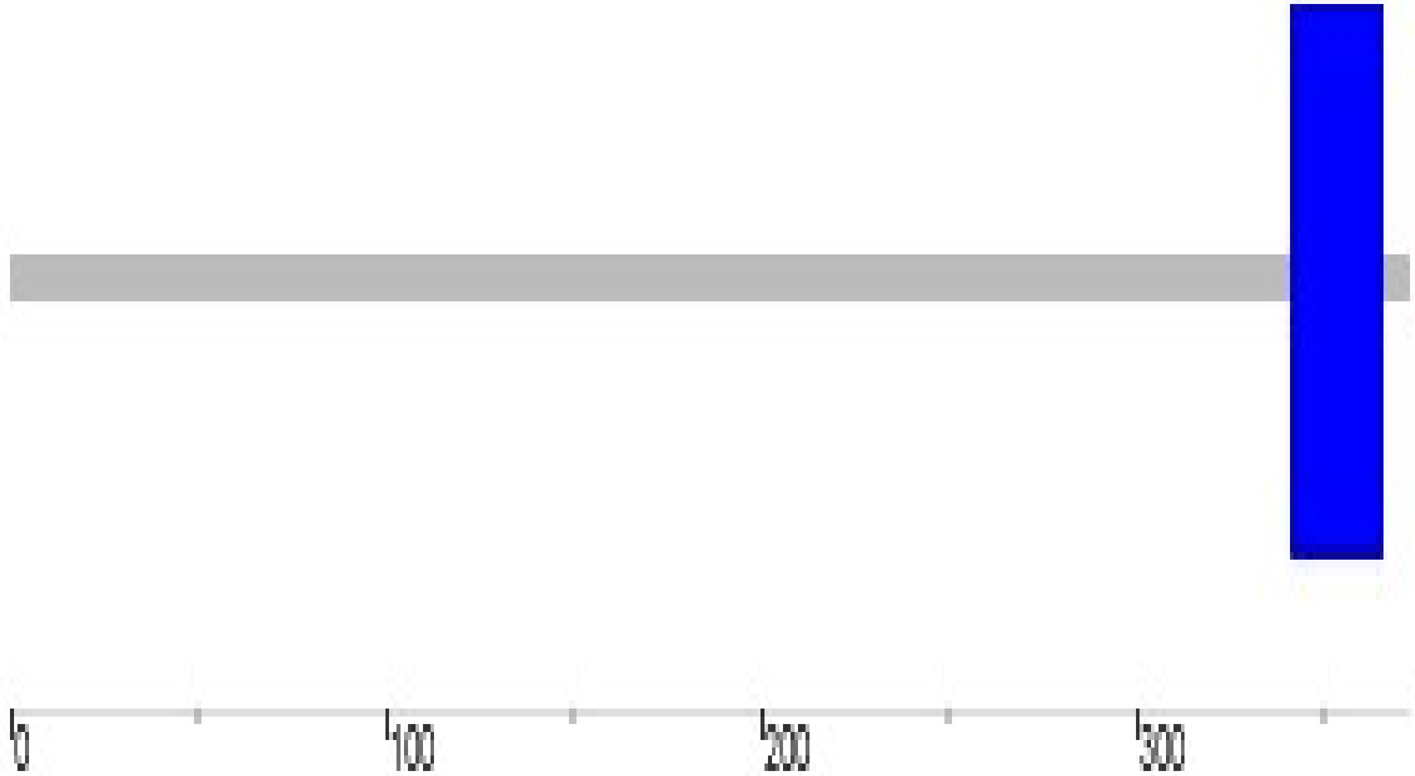
跨膜区域分析


ProtScale output for user sequence
Transmembrane tendency
score
1.5
1
0.5
0
-0.5
-1
-1.5
-2
50
100
150
200
250
300
350
Position

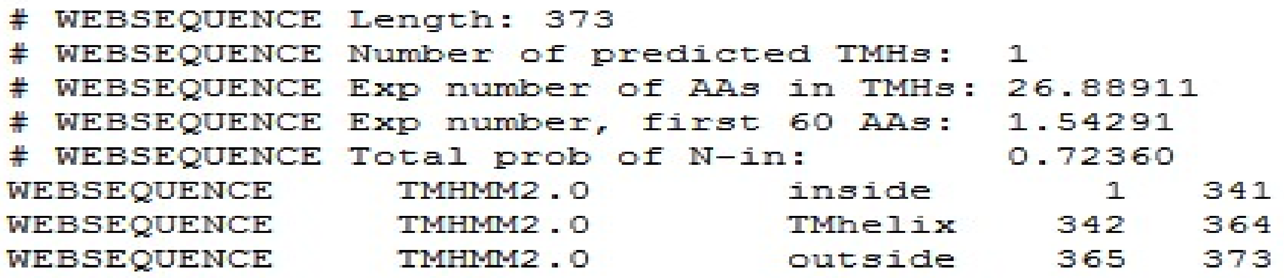
Weblab Tmap 跨膜分析



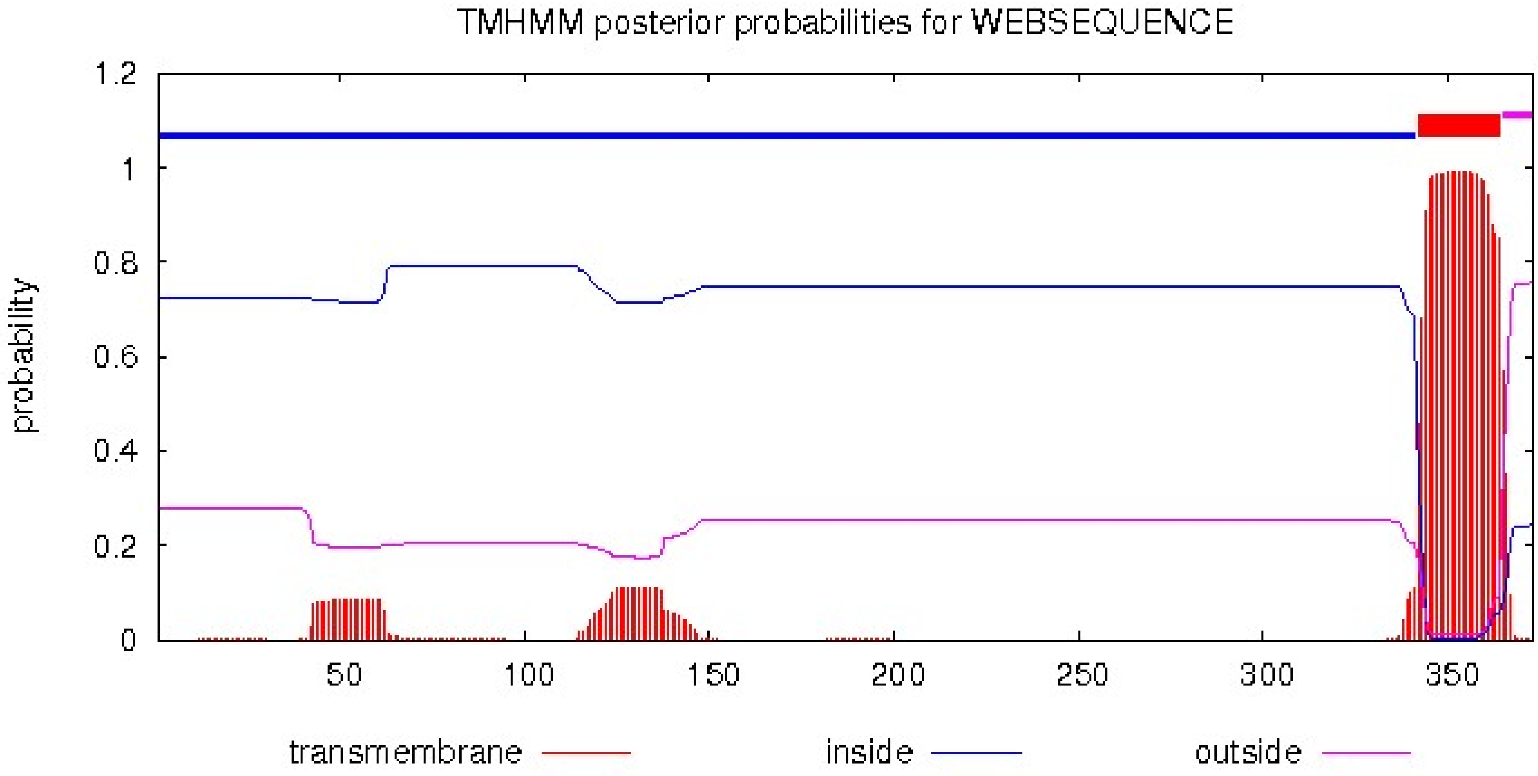
SMART 分析



TMHMM分析



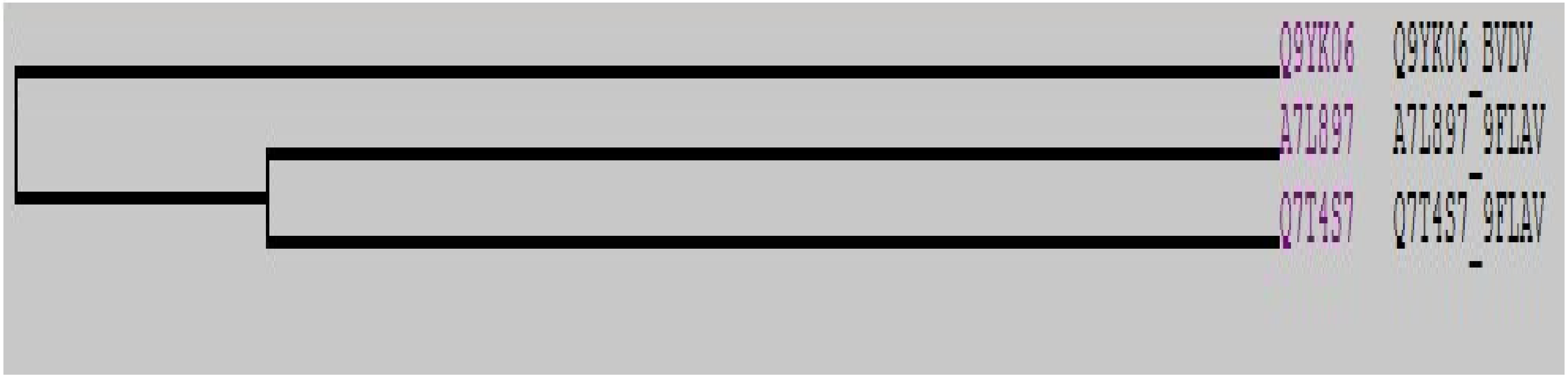
```
# WEBSEQUENCE Length: 373
# WEBSEQUENCE Number of predicted TMHs:  1
# WEBSEQUENCE Exp number of AAs in TMHs: 26.88911
# WEBSEQUENCE Exp number, first 60 AAs:  1.54291
# WEBSEQUENCE Total prob of N-in:        0.72360
WEBSEQUENCE     TMHMM2.0        inside       1   341
WEBSEQUENCE     TMHMM2.0        TMhelix    342   364
WEBSEQUENCE     TMHMM2.0        outside    365   373
```

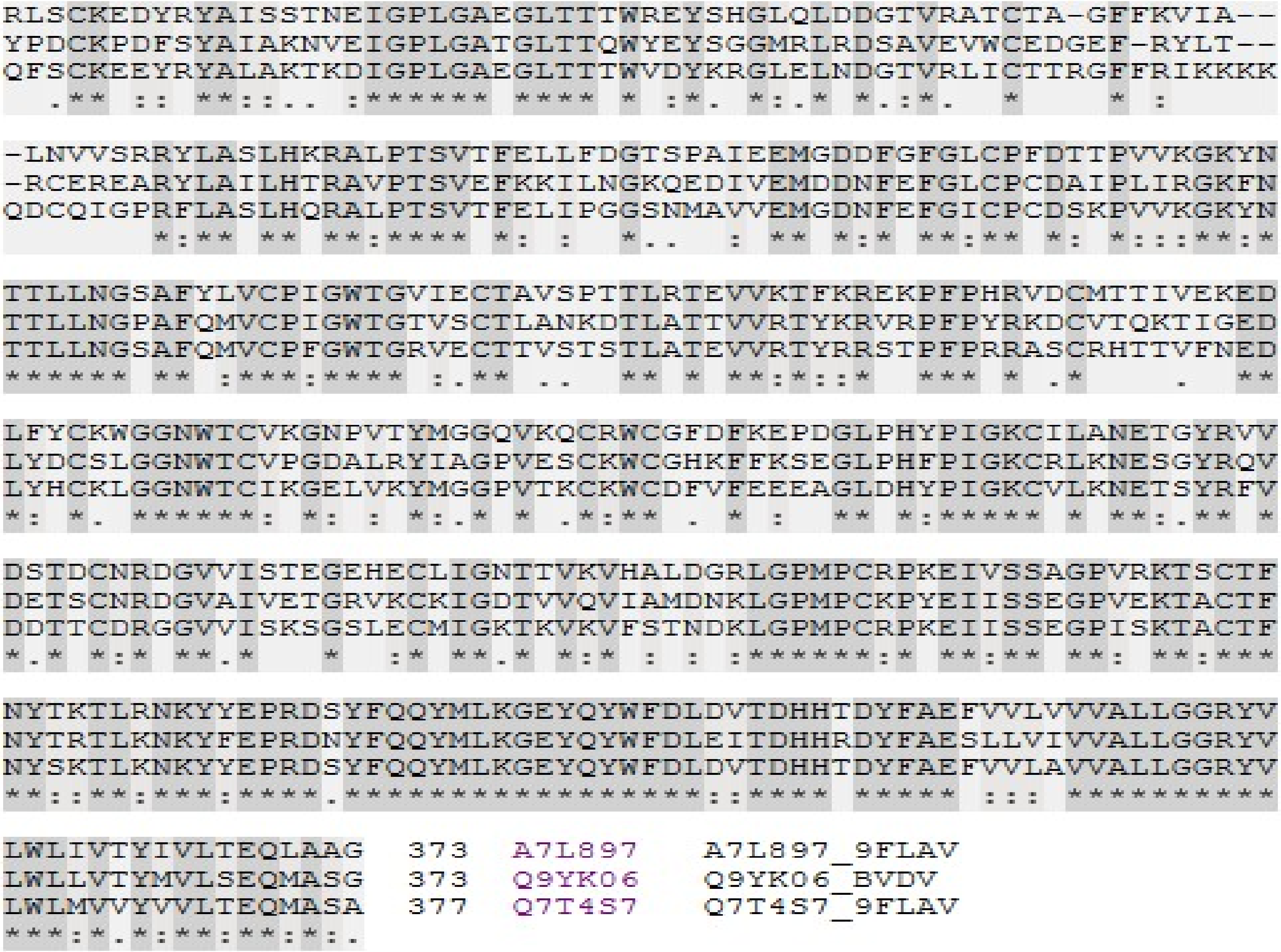


TMHMM posterior probabilities for WEBSEQUENCE

瘟病毒属毒株序列比对

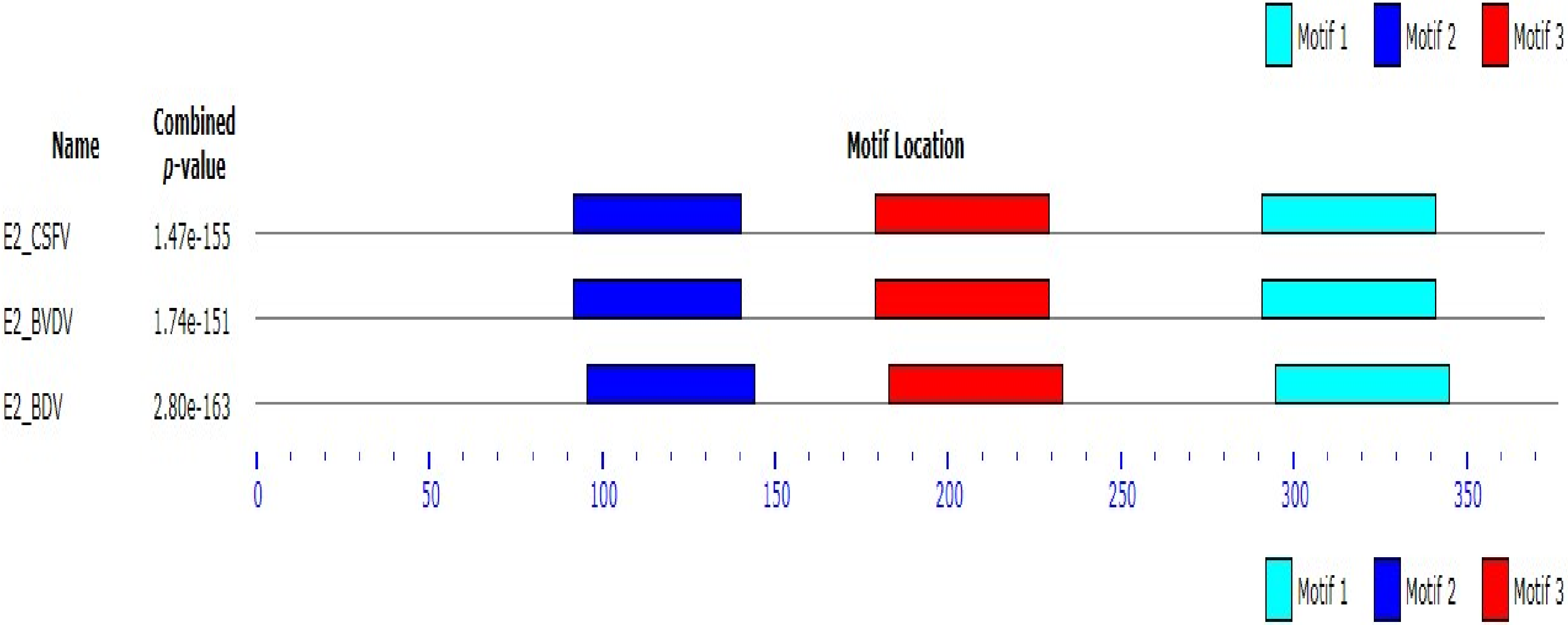
Identical positions	198
Identity	52.52%
Similar positions	85



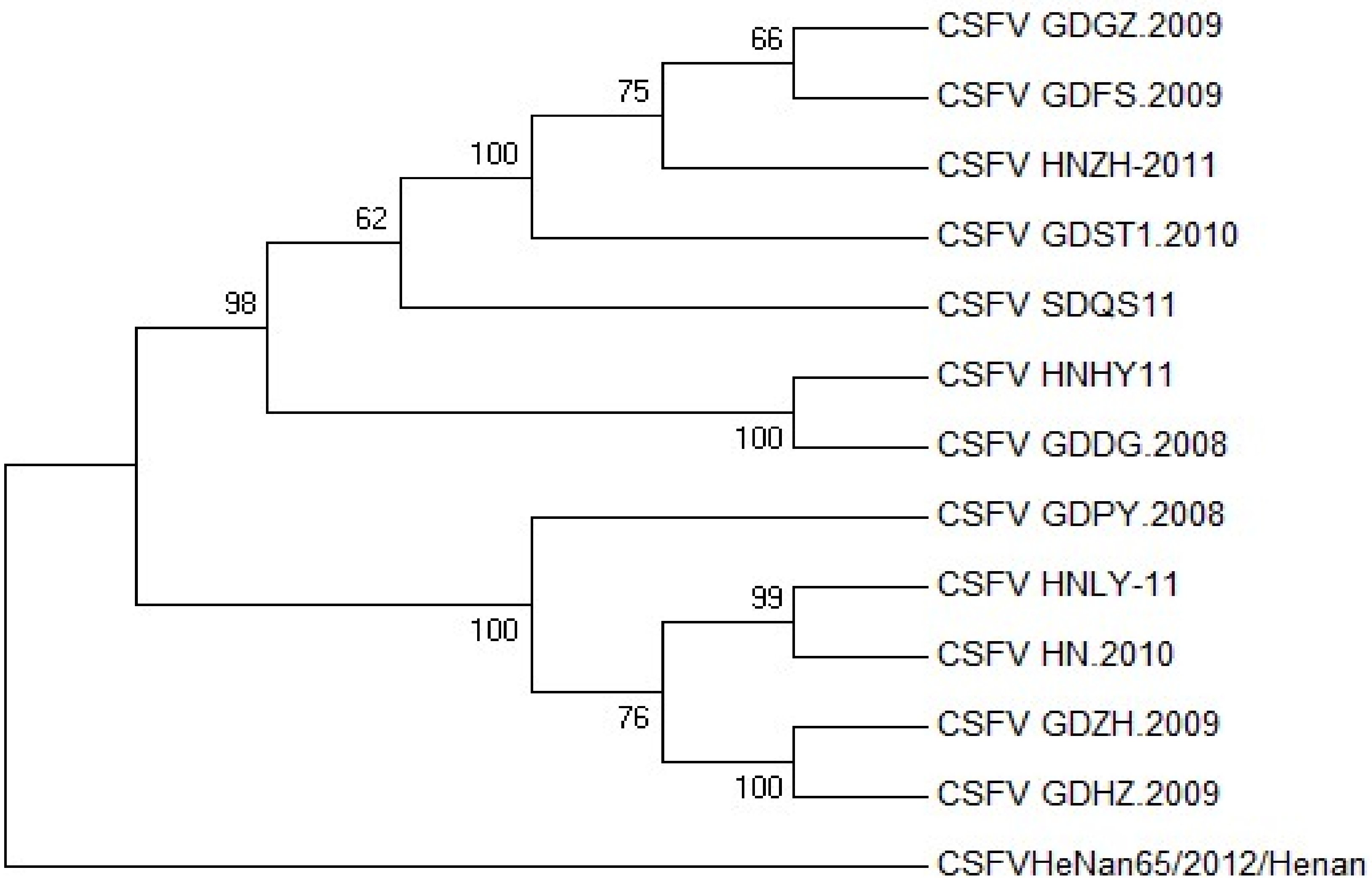


```
RLSCKEDYRYAISSTNEIGPLGAEGLTTTWREYSHGLQLDDGTVRATCTA-GFFKVIA--
YPDCKPDFSYAIAKNVEIGPLGATGLTTQWYEYSGGMRLRDSAVEVWCEDGEF-RYLT--
QFSCKEEYRYALAKTKDIGPLGAEGLTTTWVDYKRGLELNDGTVRLICTTRGFFRIKKKK
  .** :: **::.. :****** **** * :*. *:.* *.:*.  *    * :     

-LNVVSRRYLASLHKRALPTSVTFELLFDGTSPAIEEMGDDFGFGLCPFDTTPVVKGKYN
-RCEREARYLAILHTRAVPTSVEFKKILNGKQEDIVEMDDNFEFGLCPCDAIPLIRGKFN
QDCQIGPRFLASLHQRALPTSVTFELIPGGSNMAVVEMGDNFEFGICPCDSKPVVKGKYN
       *:** ** **:**** *: :  *..  : ** *:* **:** *: *:::**:*

TTLLNGSAFYLVCPIGWTGVIECTAVSPTTLRTEVVKTFKREKPFPHRVDCMTTIVEKED
TTLLNGPAFQMVCPIGWTGTVSCTLANKDTLATTVVRTYKRVRPFPYRKDCVTQKTIGED
TTLLNGSAFQMVCPFGWTGRVECTTVSTSTLATEVVRTYRRSTPFPRRASCRHTTVFNED
****** ** :***:**** :.** ..  ** * **:*::*  *** * .*    .  **

LFYCKWGGNWTCVKGNPVTYMGGQVKQCRWCGFDFKEPDGLPHYPIGKCILANETGYRVV
LYDCSLGGNWTCVPGDALRYIAGPVESCKWCGHKFFKSEGLPHFPIGKCRLKNESGYRQV
LYHCKLGGNWTCIKGELVKYMGGPVTKCKWCDFVFEEEAGLDHYPIGKCVLKNETSYRFV
*: *. ******: *: : *:.* * .*:** . * :  ** *:***** * **:.** *

DSTDCNRDGVVISTEGEHECLIGNTTVKVHALDGRLGPMPCRPKEIVSSAGPVRKTSCTF
DETSCNRDGVAIVETGRVKCKIGDTVVQVIAMDNKLGPMPCKPYEIISSEGPVEKTACTF
DDTTCDRGGVVISKSGSLECMIGKTKVKVFSTNDKLGPMPCRPKEIISSEGPISKTACTF
*.* *:* **.*   *  :* **.* *:* : : :******:* **:** **: **:***

NYTKTLRNKYYEPRDSYFQQYMLKGEYQYWFDLDVTDHHTDYFAEFVVLVVVALLGGRYV
NYTRTLKNKYFEPRDNYFQQYMLKGEYQYWFDLEITDHHRDYFAESLLVIVVALLGGRYV
NYSKTLKNKYYEPRDSYFQQYMLKGEYQYWFDLDVTDHHTDYFAEFVVLAVVALLGGRYV
**::**:***:****.*****************::**** ***** ::: **********

LWLIVTYIVLTEQLAAG  373  A7L897   A7L897_9FLAV
LWLLVTYMVLSEQMASG  373  Q9YK06   Q9YK06_BVDV
LWLMVVYVVLTEQMASA  377  Q7T4S7   Q7T4S7_9FLAV
***:*.*:**:**:*:.
```

MEME分析保守基序


Motif 1
Motif 2
Motif 3
Name
Combined p-value
Motif Location
E2_CSFV
1.47e-155
E2_BVDV
1.74e-151
E2_BDV
2.80e-163
0
50
100
150
200
250
300
350
Motif 1
Motif 2
Motif 3

序列比对


66
75
100
62
98
100
100
99
76
100
CSFV GDGZ.2009
CSFV GDFS.2009
CSFV HNZH-2011
CSFV GDST1.2010
CSFV SDQS11
CSFV HNHY11
CSFV GDDG.2008
CSFV GDPY.2008
CSFV HNLY-11
CSFV HN.2010
CSFV GDZH.2009
CSFV GDHZ.2009
CSFVHeNan65/2012/Henan

谢谢