

CEA及其实例 (CEAM5) 分析

王群 孟艳艳 王圣惠 李世贵 刘少卿

2008年1月13日

报告内容

- CEA简介
 - CEACAM基本序列分析
 - 实例（CEAM5）分析
-

CEA简介

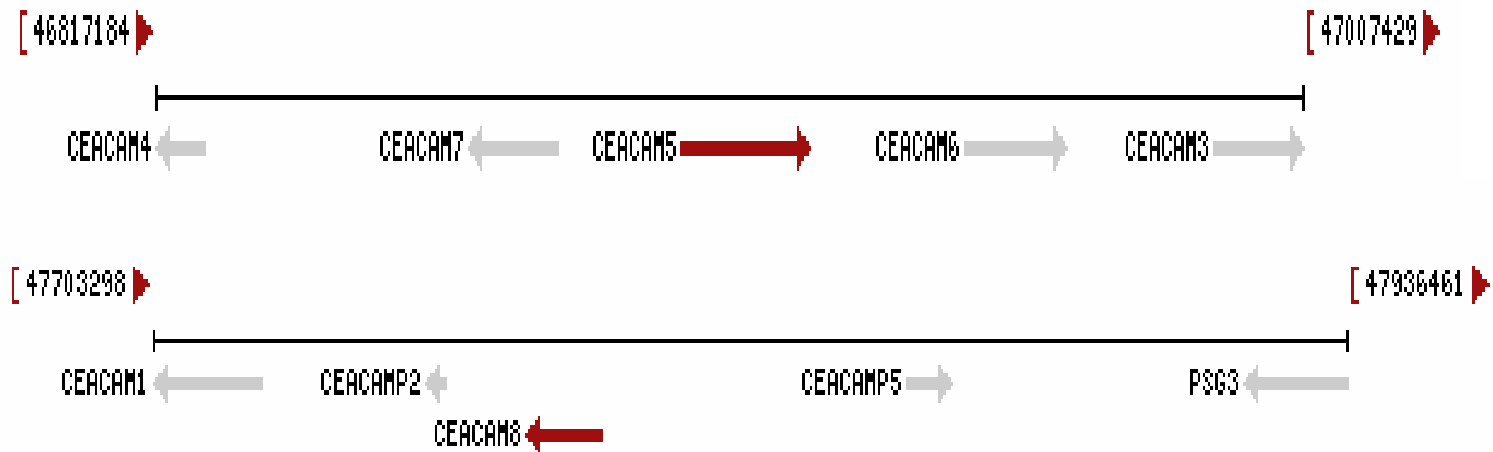
- Carcinoembryonic antigen, CEA癌胚抗原
 - 位于第19号染色体的长臂端
 - 29个基因串联
 - 两个亚家族
-

两个亚家族

- CEA-related cell adhesion molecule, CEACAM
癌胚抗原相关细胞黏附分子
 - pregnancy specific-glycoprotein, PSG
妊娠特异性糖蛋白
-

CEACAM基本序列分析

- 属于免疫球蛋白超家族
- 具有1-7个免疫球蛋白样结构域



实例 (CEAM5) 分析

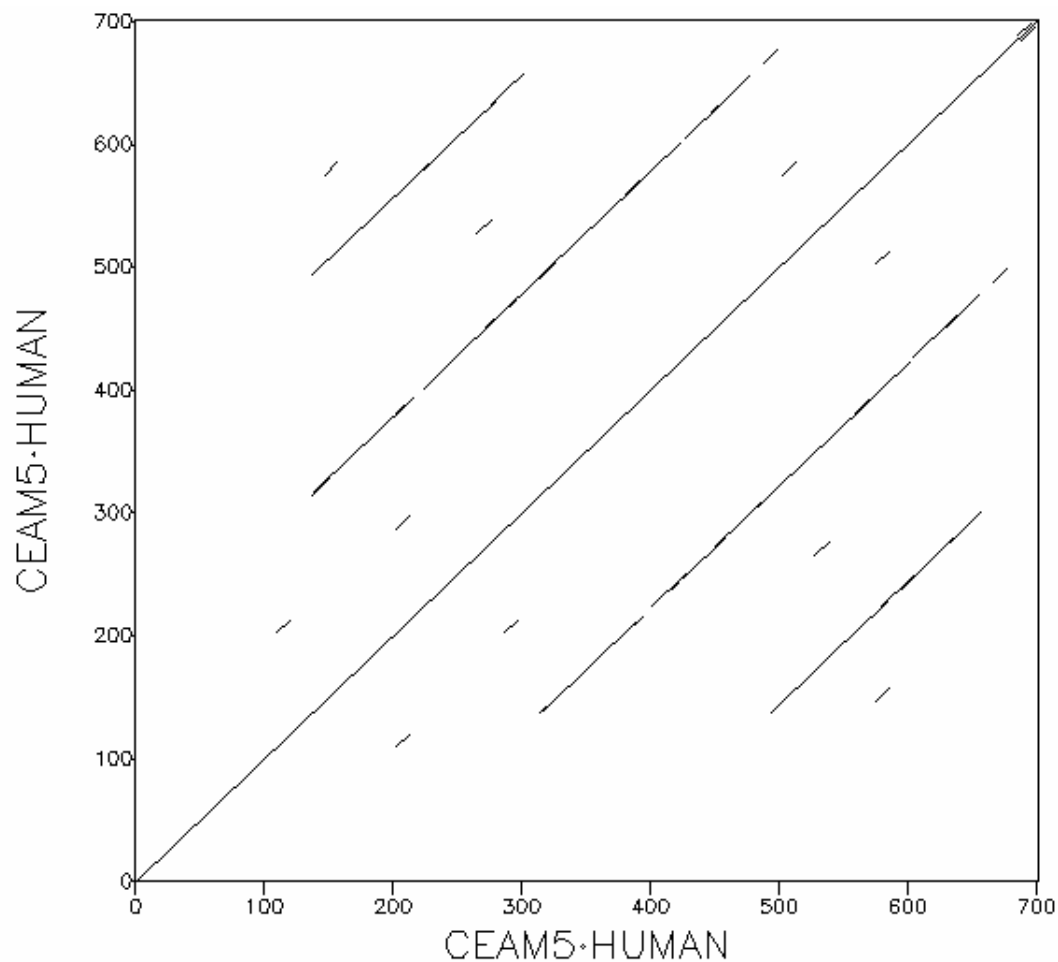
- 通常所说的**CEA** 主要指**CEAM5**
- 肿瘤细胞抗原
- **9**个外显子
- 成熟体有**668**个氨基酸



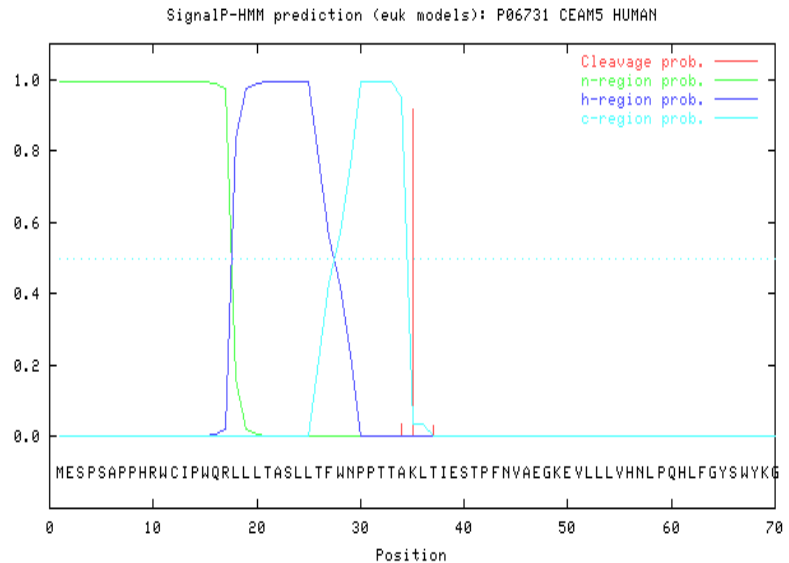
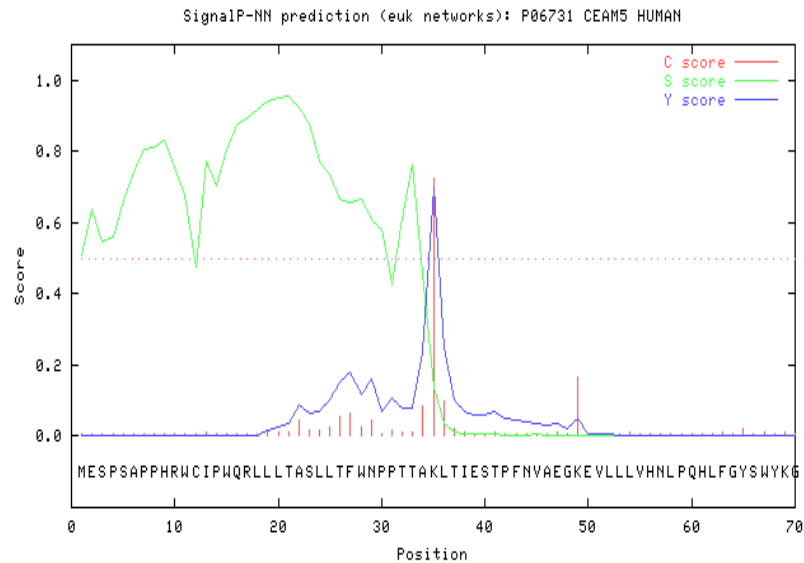
CEAM5分析

- **1、Dotmatcher序列比对**
 - **2、信号肽预测**
 - **3、跨膜分析**
 - **4、糖基化位点预测**
 - **5、保守序列分析**
 - **6、结构功能域预测**
 - **7、3D结构预测**
-

1、Dotmatcher序列比对



2、信号肽(SignalP)分析



```
>P06731_CEAM5_HUMAN      length = 70
# Measure   Position   Value   Cutoff   signal peptide?
max. C      35         0.723   0.32    YES
max. Y      35         0.699   0.33    YES
max. S      21         0.957   0.87    YES
mean S      1-34       0.723   0.48    YES
           D      1-34       0.711   0.43    YES
# Most likely cleavage site between pos. 34 and 35: TTA-KL
```

3、跨膜分析

```

Protein: noname
Length: 702
N-terminus: OUT
Number of transmembrane helices: 1
Transmembrane helices: 683-701

Total entropy of the model: 17.0191
Entropy of the best path: 17.0199

The best path:

seq MESFSAPFHR WCIFWQRLLL TASLLTFWNP FTTAKLTIES TFFNVAEGKE 50
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq VLLLVHNLFP HLFGYSWYKG ERVDGNRQII GYVIGTQQAT PGPAYSGREI 100
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq IYFNASLLIQ NIIQNDFGY TLHVIRKSDLV NEAATQFRV YBELFKFSIS 150
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq SNNSKFVEDK DAVAFTCEFE IQDATYLNWV NNQSLFVSPR LQLSNGNRIL 200
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq TLFNVTNRMT ASYKQETQNP VSARRSDSVI LNVLYGPDAP TISPLNTSYR 250
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

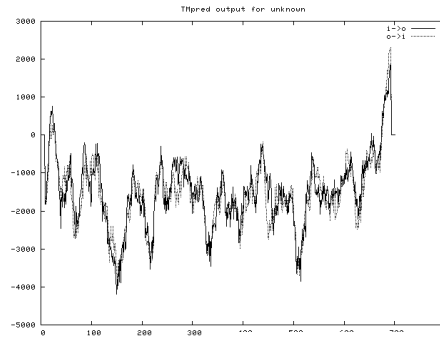
seq SGENLNLSCH AASNPPAQYS WFNVTGQQS TQELFIPNIT VNNSGSYTCQ 300
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq AHNSTGLNLR TIVTITVVA EPPKPIITSN NSNPEVEDA VALTCEPPIQ 350
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq NITYLWVNVN QSLFVSPRLQ LSNDRILTL LSVIRNDVGP YECGIQNELS 400
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO

seq VHSDFVILN VLYGDDPTII SPSVTYRFG VNLSLSCHAA SNPPAQYSWL 450
pred OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO
    
```

HMMTOP预测结果



TMpred output的预测结果

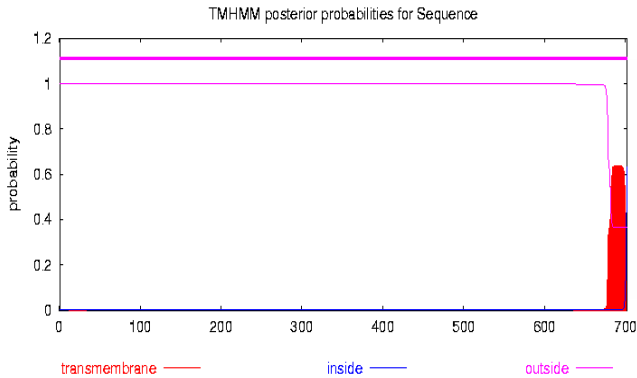
The sequence positions in brackets denominate the core region. Only scores above 500 are considered significant.

Inside to outside helices : 3 found-

from	to	score	center
11 (15)	34 (34)	762	24.
645 (647)	669 (663)	40	655.
683 (683)	702 (702)	1858	692.

Outside to inside helices : 2 found-

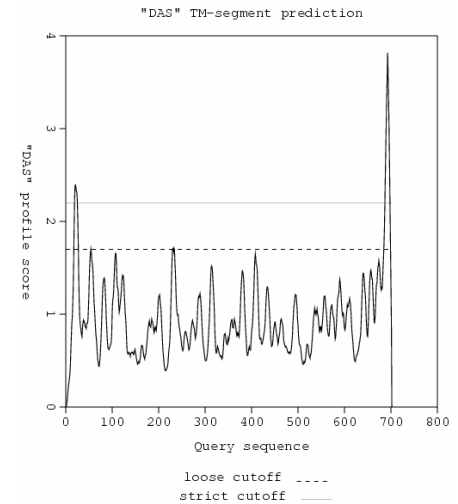
from	to	score	center
13 (17)	36 (36)	294	28.
685 (685)	702 (702)	2298	693.



TMHMM预测结果

```

# Sequence Length: 702
# Sequence Number of predicted TMHs: 0
# Sequence Exp number of AAs in TMHs: 13.64783
# Sequence Exp number, first 60 AAs: 0.0229
# Sequence Total prob of N-in: 0.00150
Sequence      TMHMM2.0      outside      1      702
    
```



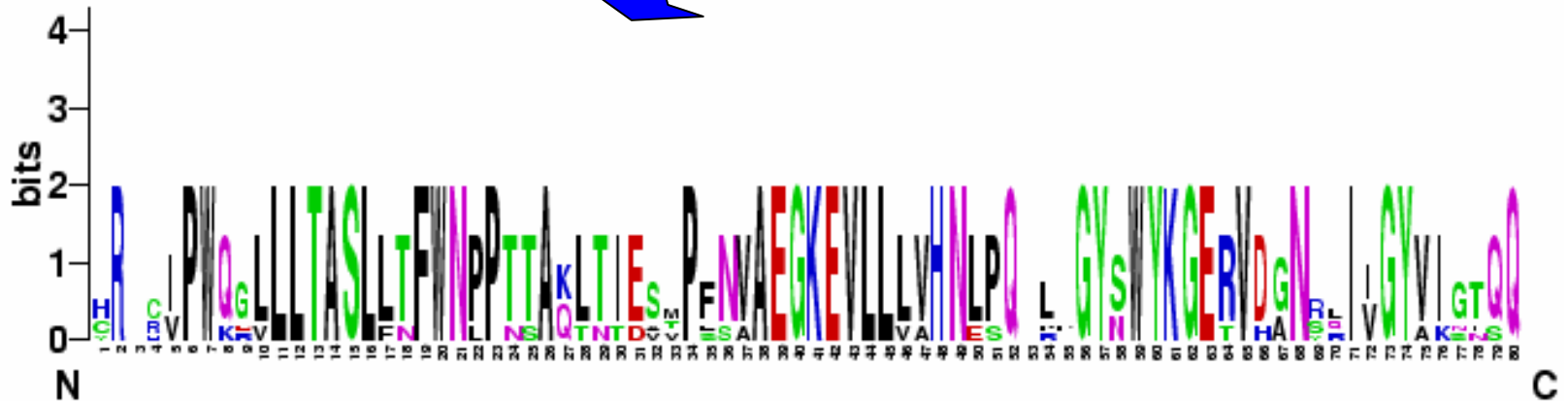
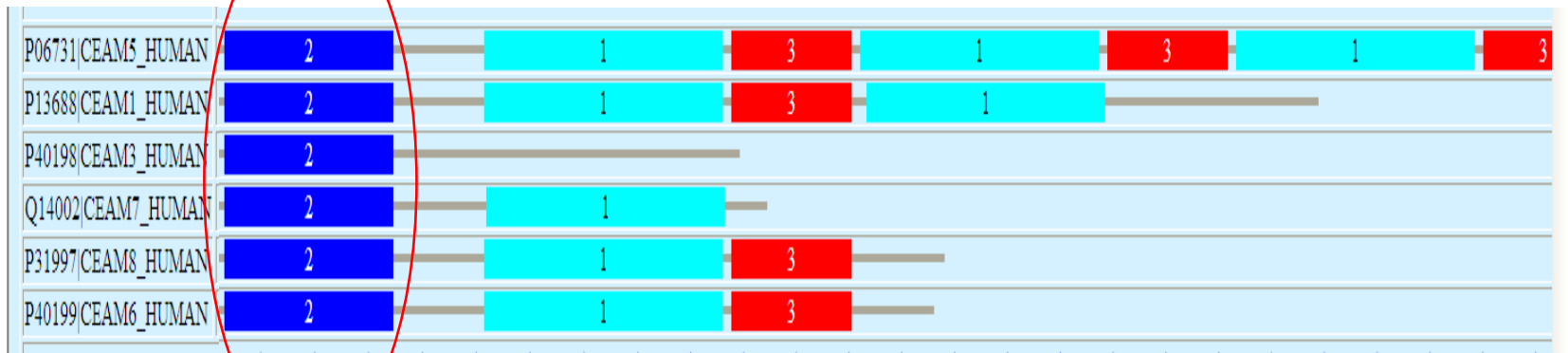
DAS预测结果

4、糖基化位点预测

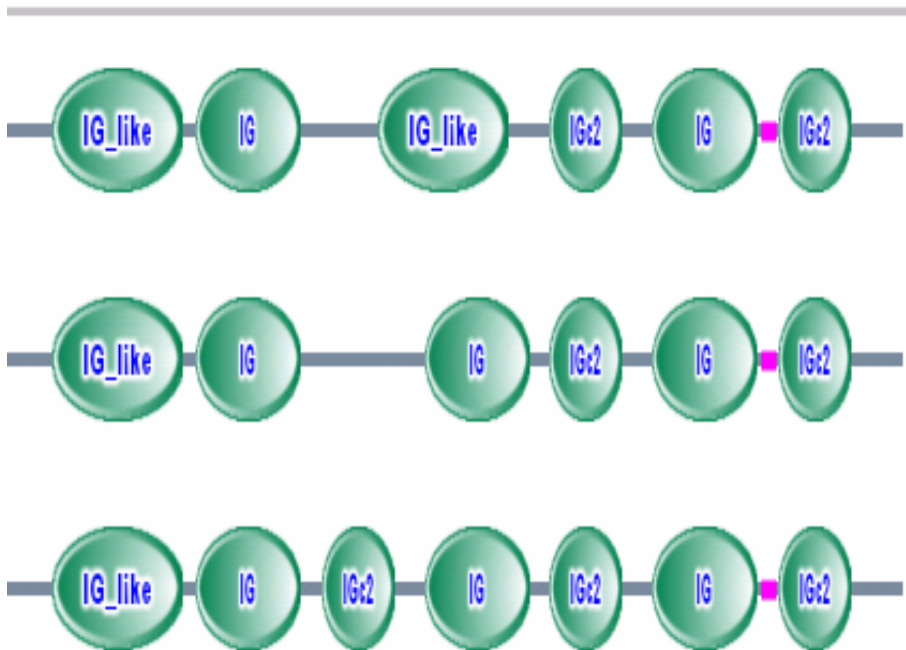
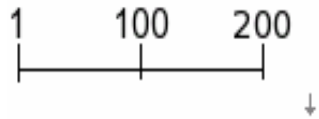
Name: P06731_CEAM5_HUMAN	Length: 702	
MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTAKLTIESTPFNVAEGKEVLLLVDHNLPHQLFGYSWYKGERVDGHRQII		80
GYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDK		160
DAVAFTCEPETQDATYLWVWNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYK CETQNPVSARRSDSVILNVLYGPDAP		240
TISPLNTSYRSGENLHLSCHAASNPPAQYSWVNGTFQQSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTTVTTITVYA		320
EPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWVWNNQSLPVSPRLQLSNDNRTL TLLSVTRNDVGPYECGIQNELS		400
VDHSDPVILNVLYGPDPTISPSYTYRPGVNL SLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQAN		480
NSASGHSRTTVKTI TVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWVWNGQSLPVSPRLQLSNGNRTLTLFN		560
VTRNDARAYVCGIQNSVSA NRS DPVTL DVLYGPDPTIISPDDSSYLSGANLHLSCHSASNPSPOYSWRINGIPQQHTQVL		640
FIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGMIGVLVGVALI		702
.....		80
.....N.....N.....N.....		160
.....N.....N.....N.....N.....		240
.....N.....N.....N.....N.....N.....N.....		320
.....N.....N.....N.....N.....		400
.....N.....N.....N.....		480
.....N.....N.....N.....		560
.....N.....N.....		640
.....N.....		720

Asn-Xaa-Ser/Thr序列（蓝色） 糖基化的Asn位点（红色）

5、保守结构域分析



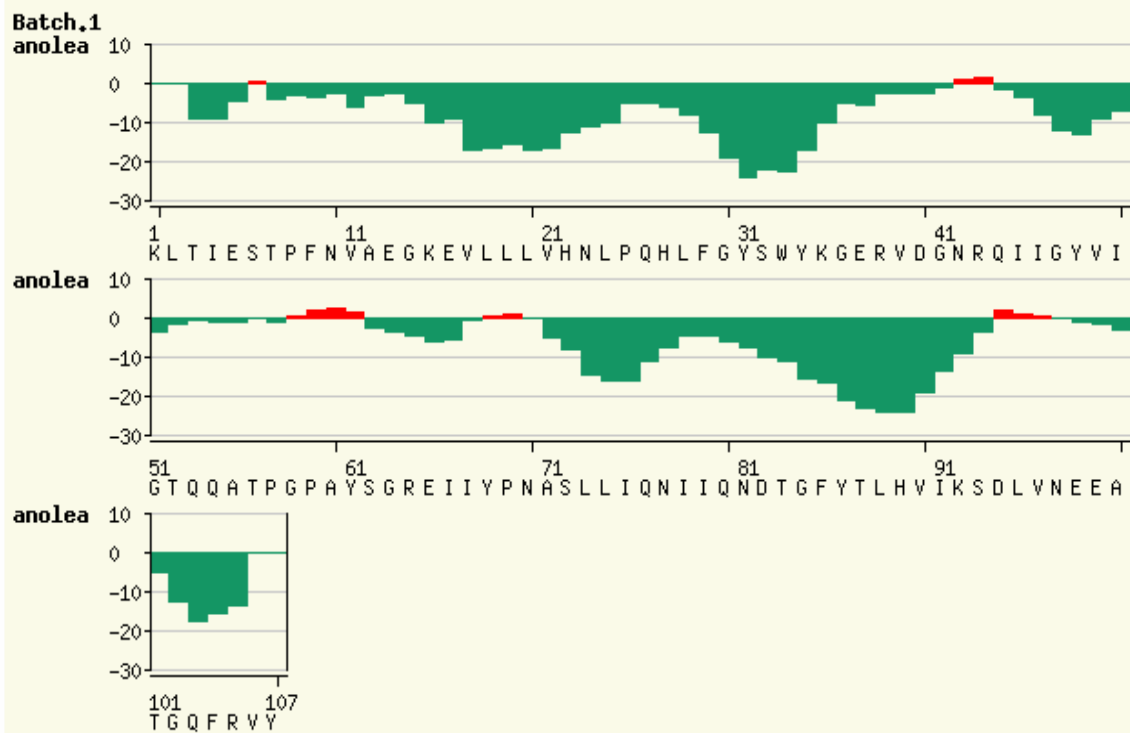
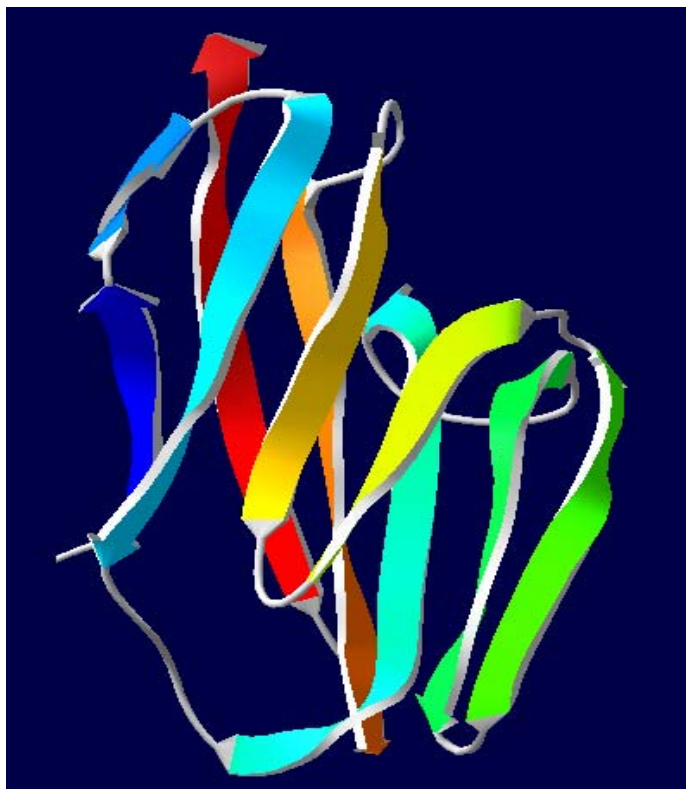
6、SMART结构功能域预测



Confidently predicted domains, repeats, motifs and features:

Name	Begin	End	E-value
IG_like	40	141	6.35e+00
IG	152	234	2.37e-05
IGc2	250	306	6.65e-09
IG_like	294	395	6.15e-01
IG	330	412	2.64e-03
IGc2	428	484	3.56e-09
IG	508	590	4.36e-04
low complexity	593	604	-
IGc2	606	662	9.49e-05

7、Swiss Model 3D结构预测



谢谢大家
请多指教
祝大家春节愉快!

王群 孟艳艳 王圣惠 李世贵 刘少卿