

bioinformatics analysis and epitope prediction of EIAV gp90

马传染性贫血病毒囊膜糖蛋白gp90 的生物信息学分析及抗原表位预测

哈尔滨兽医研究所

第**G19**组
报告人：王春花

组长：陈玲凤
组员：单继艳 孙留克 王春花

目 录

- ◆ 课题背景
- ◆ 生物信息学分析
- ◆ 抗原表位预测

课题背景

马传染性贫血病（EIA）是由逆转录病毒科慢病毒属的马传染性贫血病毒（EIAV）引起的马属动物传染病。以重复发生的病毒血症以及发热,贫血,水肿,血小板减少症和消瘦等临床症状为特点。我国科学家自主研发的EIAV弱毒疫苗是世界上惟一大规模成功应用的慢病毒疫苗,可以同时活化T细胞免疫与体液免疫。EIAV与HIV同为慢病毒,两者在病毒形态、基因组结构、细胞嗜性、病毒复制的分子机制、病毒的生活周期、抗原漂移规律、免疫机制及病毒与宿主相互作用等方面都极为相似。深入探索EIAV减毒疫苗相关的免疫保护机制,将为HIV的新型疫苗的研究提供借鉴。

GP90是EIAV env基因编码的糖蛋白,它能够通过和细胞受体的相互作用而影响细胞的嗜性,也是介导细胞融合和病毒进入细胞过程中的重要作用成分,此外还是中和抗体和细胞毒T细胞的主要靶分子。Gp90是个高度糖基化的蛋白质,是诱导宿主产生中和抗体的主要抗原

生物信息学分析

1. gp90理化性质分析
2. 二级结构预测
3. 跨膜螺旋预测
4. gp90的柔韧性、表面可及性与亲水性的预测
5. 信号肽切割位点预测
6. 蛋白跨膜螺旋的pepwheel预测
7. gp90的空间结构预测

1. gp90理化性质分析

运用在线的网络服务器<http://www.expasy.org/>分析该糖蛋白分子量、等电点、氨基酸组成等情况

理化性质分析（使用ProtParam），结果如下：

Number of amino acids: 445

Molecular weight: 50621.9

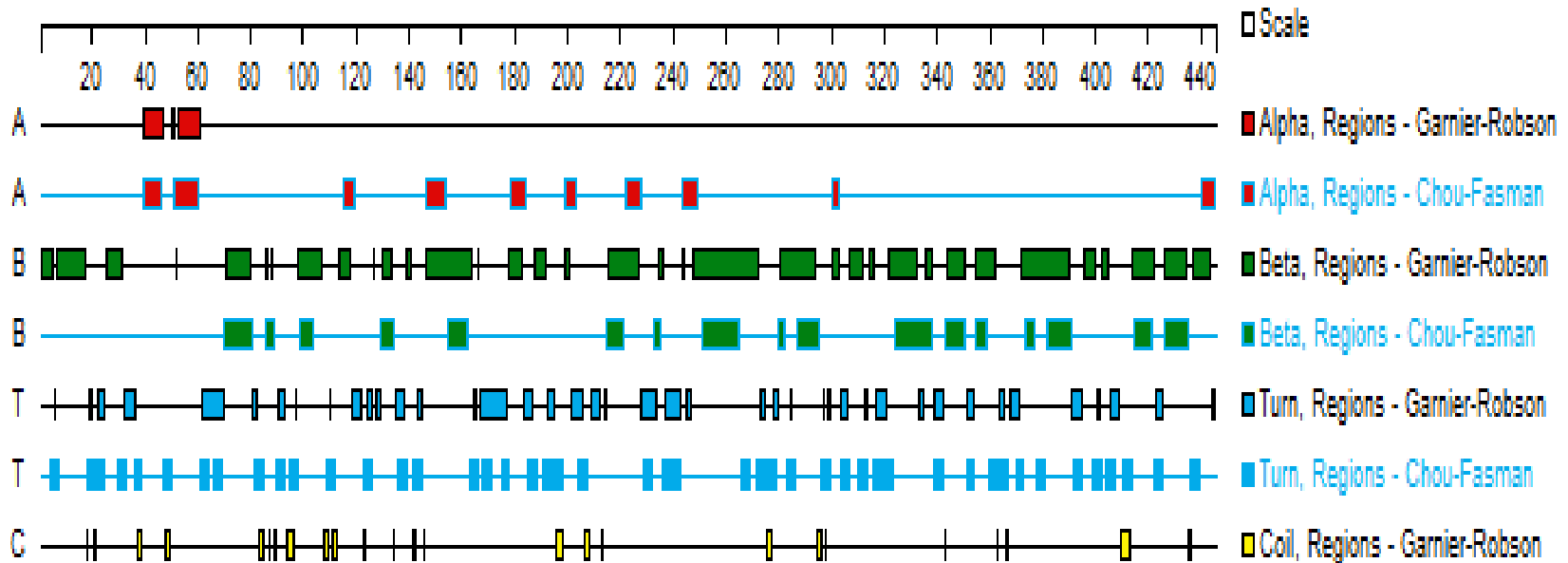
Theoretical pI: 8.03

氨基酸 ₁	比列 ₁	氨基酸 ₁	比列 ₁	氨基酸 ₁	比列 ₁	氨基酸 ₁	比列 ₁
Ala(A) ₁	3.4% ₁	Gln(Q) ₁	5.2% ₁	Leu(L) ₁	6.7% ₁	Ser(S) ₁	7.9% ₁
Arg(R) ₁	5.2% ₁	Glu(E) ₁	5.2% ₁	Lys(K) ₁	4.5% ₁	Thr(T) ₁	7.6% ₁
Asn(N) ₁	10.1% ₁	Gly(G) ₁	6.7% ₁	Met(M) ₁	2.2% ₁	Trp(W) ₁	2.0% ₁
Asp(D) ₁	3.8% ₁	His(H) ₁	2.5% ₁	Phe(F) ₁	3.4% ₁	Tyr(Y) ₁	4.3% ₁
Cys(C) ₁	4.5% ₁	Ile(I) ₁	6.5% ₁	Pro(P) ₁	3.8% ₁	Val(V) ₁	4.5% ₁

Gp90是个高度糖基化的蛋白质，是诱导宿主产生中和抗体的主要抗原，其中和表位的特异性与糖侧链的位置密切相关，抗原表位依赖于糖侧链的存在，同时也可以被糖侧链所掩盖。其中糖侧链主要是与N链接的糖基化有关

2. 二级结构预测

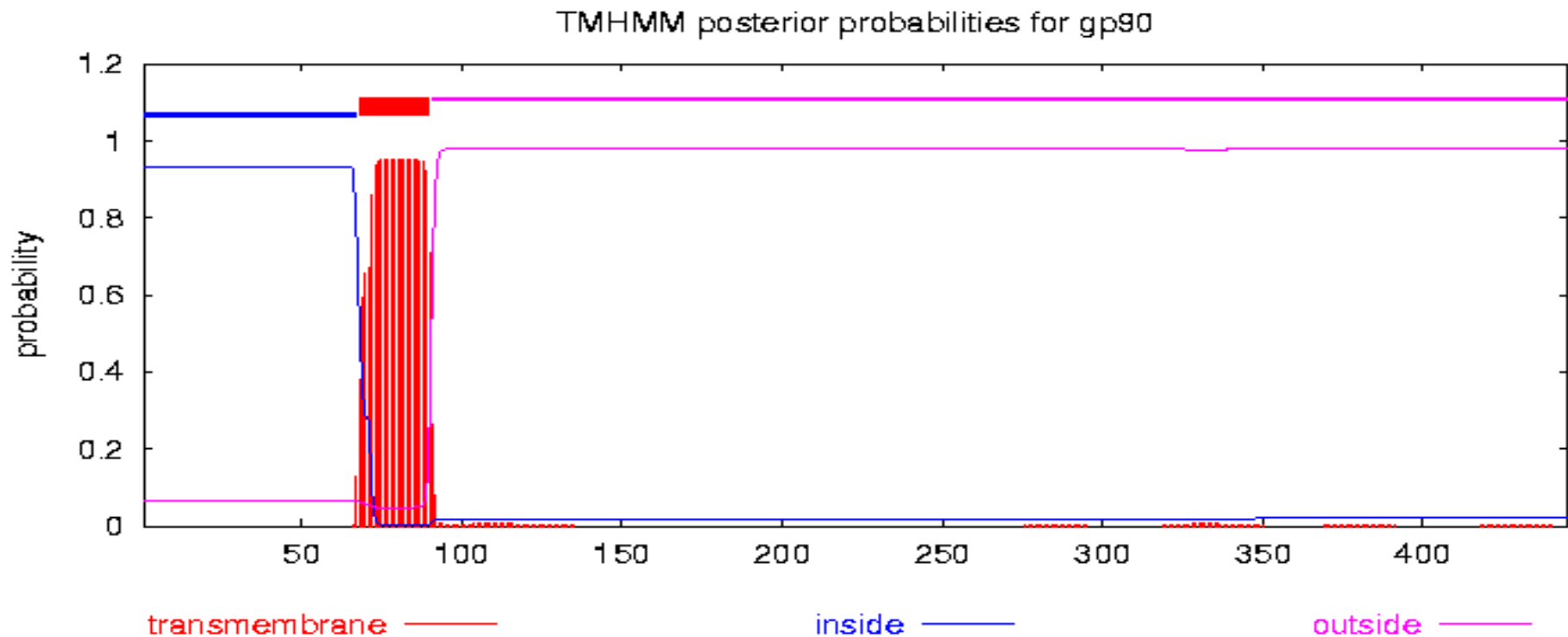
运用DNASTar中Protean软件的Chou-Fasman和Garnier-Robson方法分析gp90的 α -螺旋、 β -折叠、 β -转角和Coil卷曲所在区段



3. gp90的跨膜螺旋预测（TMHMM软件）

inside	1	67
TMhelix	68	90
outside	91	445

该蛋白可能定位在高尔基体膜，与M蛋白、N蛋白相互作用共同介导病毒的出芽

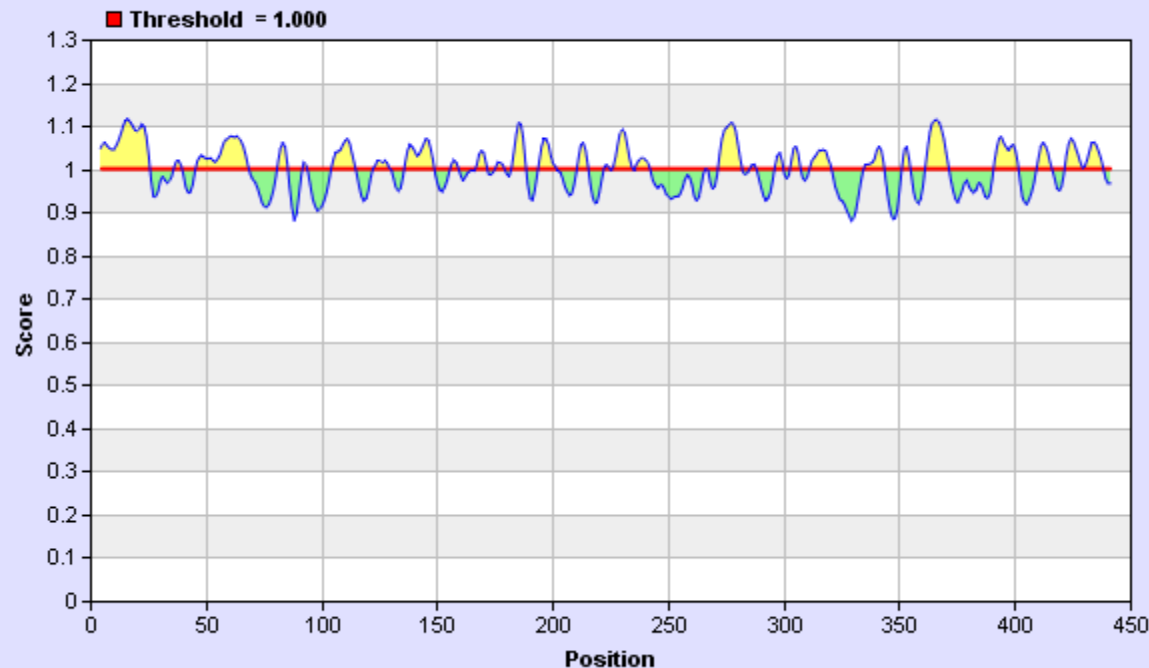


4.gp90的柔韧性、表面可及性与亲水性的预测

柔韧性预测

柔韧性参数是肽骨架自由伸展卷曲的程度，一定的柔韧性有利于与蛋白抗体结合，活动性强的氨基酸残基即柔韧性大的位点，容易形成抗原表位，在下图中（横轴为蛋白氨基酸区段，纵轴为柔韧度值）中，高于阈值基线1.0的氨基酸区段具有较好的柔韧性，是容易形成抗原表位的区域。

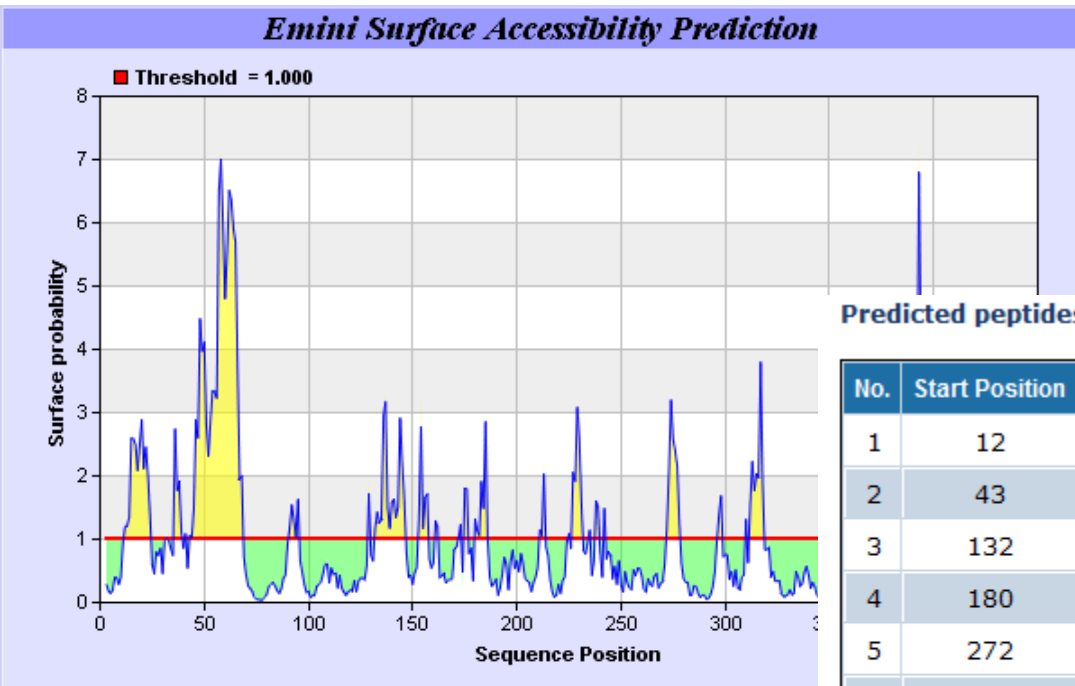
Karplus & Schulz Flexibility Prediction



4.gp90的柔韧性、表面可及性与亲水性的预测

可及性预测

可及性参数，是指蛋白质抗原中氨基酸残基被溶剂分子接触的可能性，反映了蛋白质抗原内外各层残基的分布情况，可作为判断天然蛋白中一个亚序列能形成B细胞表位的重要参数（横轴为氨基酸区段，纵轴为可及性参数数值），可及性高于基线1.0的区段易形成B细胞表位



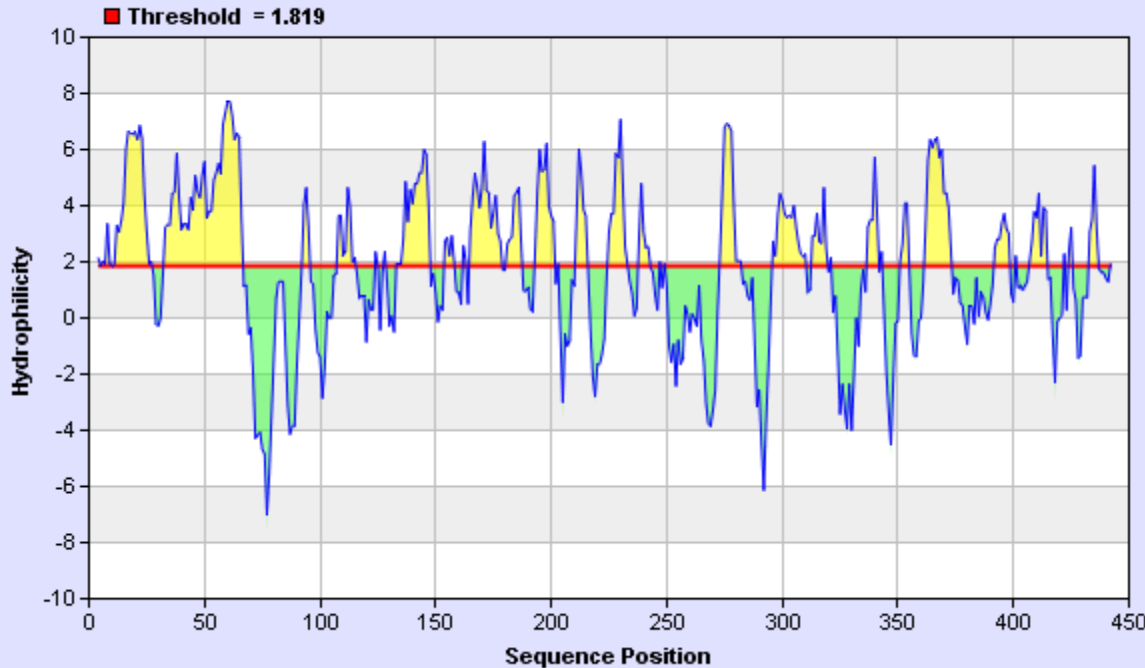
No.	Start Position	End Position	Peptide	Peptide Length
1	12	24	ITQQTESTDTQKG	13
2	43	68	AEARDTRYQEEMNRKEDKEDNKRRNN	26
3	132	146	YFSYETNRTVSRDNN	15
4	180	186	VQLKENS	7
5	272	278	KQNDTSN	7
6	312	318	TNIKRPD	7
7	363	369	SNTNSTN	7
8	391	398	PRKNKRLQ	8
9	410	416	NNKTELRL	7

4.gp90的柔韧性、表面可及性与亲水性的预测

亲水性预测

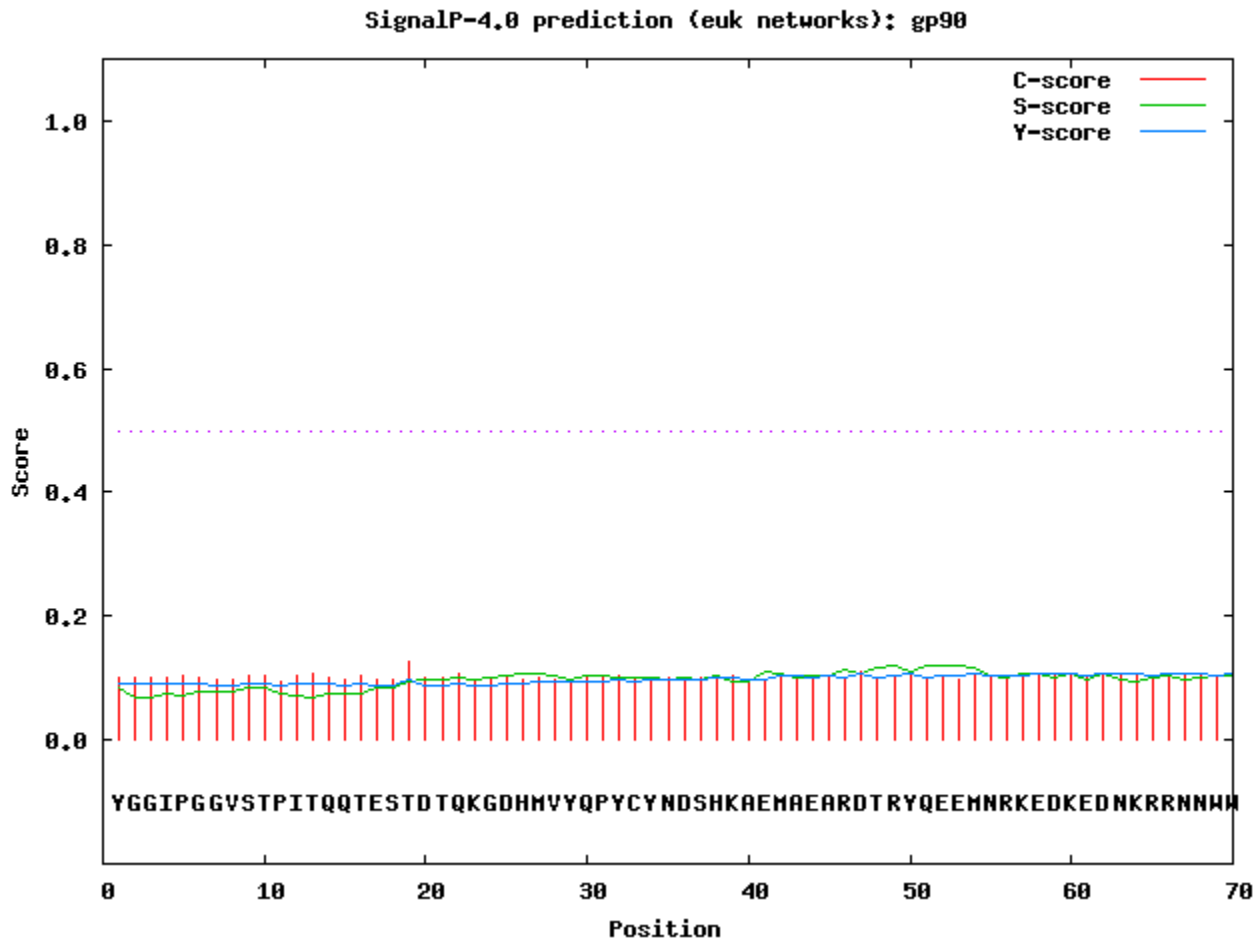
蛋白质抗原的氨基酸残基可分为亲水性残基和疏水性残基两大类。在机体内，疏水性残基一般埋在蛋白质内部，而亲水性残基位于表面，Hopp和Woods通过对蛋白质中局部亲水性（local hydrophilicity）的计算发现，局部亲水性的峰值区域大多位于线性B细胞表位内或与之紧密相连。因此蛋白质的亲水部位与蛋白抗原表位有密切的联系。由下图中（横轴为蛋白氨基酸区段，纵轴为亲水性参数值）可见，gp90蛋白具有良好的亲水性

Parker Hydrophilicity Prediction



5. 信号肽切割位点预测

信号肽是新生分泌性蛋白质的N端保守的氨基酸序列，是分泌性蛋白质所独具的。由下图可见，gp90无信号肽切割位点，也与其包膜糖蛋白的身份相符。



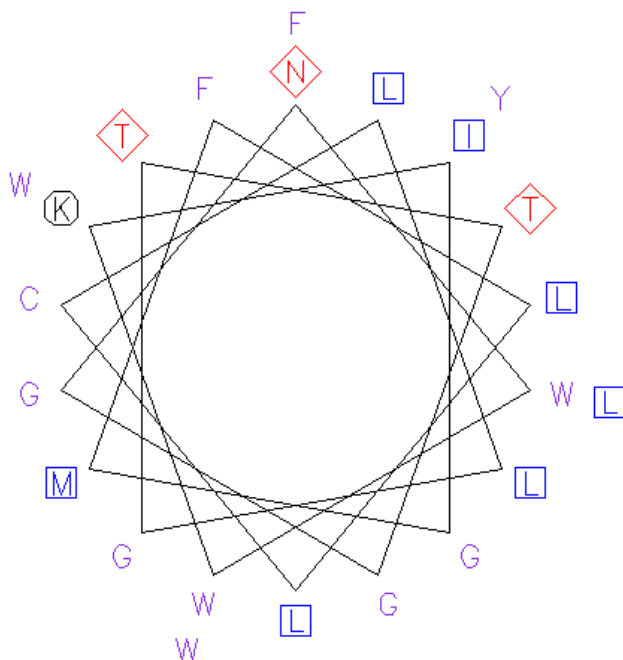
6. 蛋白跨膜螺旋的pepwheel预测

68-90

NWWKIGMFLLCLLGTTGGFLWWY

Helical wheel of fasta::541905:gp90

Tue 29 May 2012 22:42:32



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，而生物膜的主要成分是脂类，所以跨膜螺旋也应该是亲脂类的，也就是疏水的，由此看来分析结果与gp90蛋白包膜糖蛋白的身份吻合。

7. Gp90蛋白的结构分析

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

7. Gp90蛋白的结构分析（同源建模）

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI Welcome sunik. [Sign Out]

NCBI/BLAST/blastp suite Standard Protein BLAST

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

NWNKIGMFLLCLLGTTGGFLWYEGQHSHYIGLVTIGGRLNGSGMTSAIECWGSPFGCRPFTNYFS
YETNRTVSRDNNIATLLDAYQREITNIYRTSCVSDHQCQYKCKQVQLKENSNIIMNCGSSNSCEE
FRGFSWLECNQENAITILVDPDIEMQESDNNIWIPIKRCNETWARVKHCMDLLYGINRIMCVQPPF
FLFKQNDTSNNTGILSNCGELVFLGILEDNKAAVQNGSCTLHRTNIKRPDYSGFYQVPIFYICNLTG
LQSCNNGSIIISIMYESNNVQVLLCNTSNTNSINNAMVSCVQVQSGFVIGQAHVALPRNKRLQSPKF
AHYNCISINNKTELRSQVQLVKTSGITPLFISSTANTGLVRHKRD

Or, upload file [浏览...](#)

Job Title gp90
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Protein Data Bank proteins (pdb)

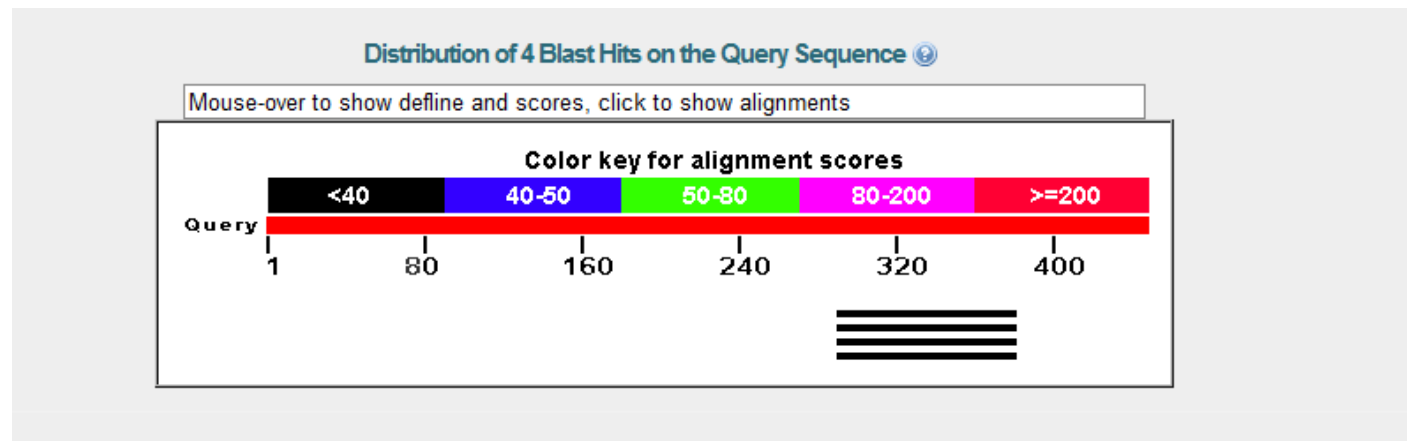
Organism Optional
Enter organism name or id—completions will be suggested Exclude [+](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional
 Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query Optional
Enter an Entrez query to limit search



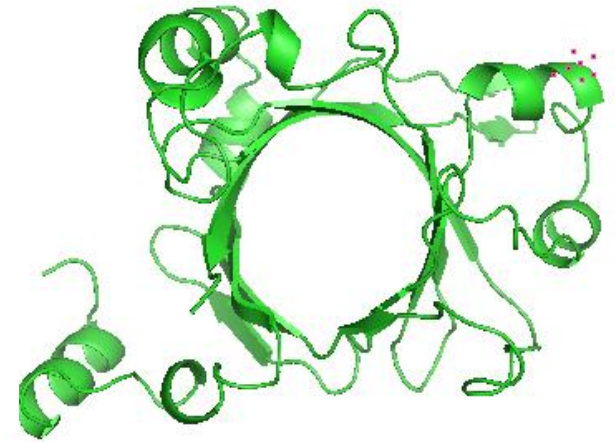
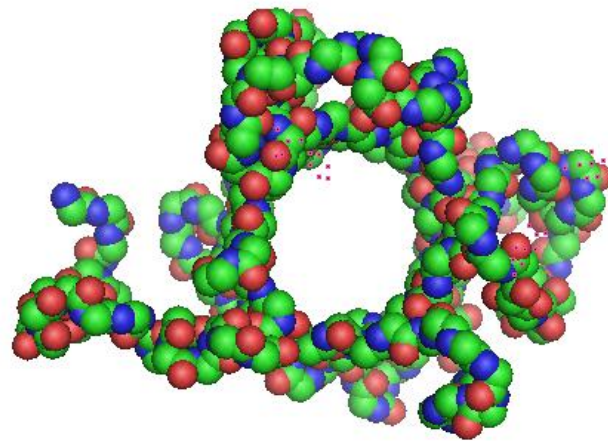
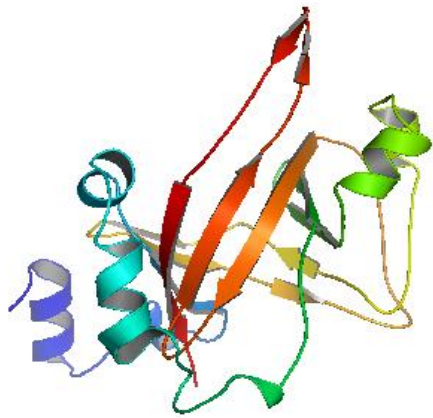
相似性太低，同源建模失败



线串法

很多一些远源的蛋白质虽然具有相似的空间结构，但其一致性小于30%，对于这类蛋白质的建模，在缺乏同源信息的情况下可选用串线法预测蛋白质折叠。该方法是将待测序列“穿”入已知蛋白质的基本骨架内，通过计算各种折叠的概率，来预测蛋白质的核心结构。

234-441



抗原表位预测

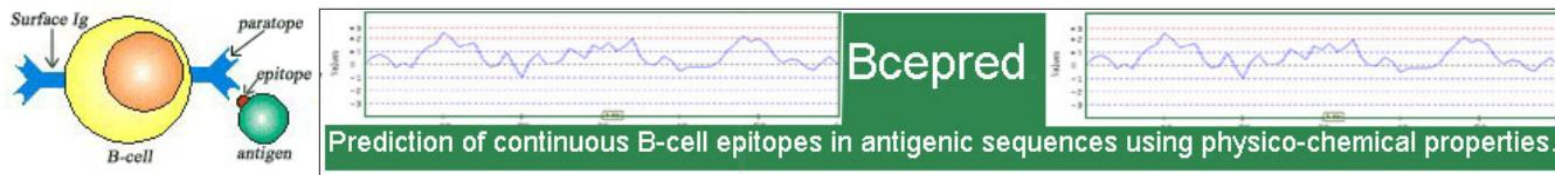
抗原表位又称抗原决定簇，是指抗原分子表面具有特殊结构和免疫活性的化学基团，具有刺激机体产生抗体或致敏淋巴细胞并能够被其识别的一个免疫活性区。根据与抗原受体结合细胞的不同，分为**B**细胞抗原表位和**T**细胞抗原表位。其中**B**细胞抗原表位则往往为亲水性肽段，一般位于抗原三维大分子的氨基酸长链折叠处，而**T**细胞表位往往仅为埋藏在蛋白质空间结构内部的疏水性肽段。

根据免疫应答的结果，可以将**B**细胞表位分为保护性表位以及毒性或限制性表位、免疫优势性表位以及亚优势性表位；根据表位的功能，又可以分为中和、非中和以及感染增强性表位、病理以及自身抗原交叉反应性表位、宿主特异性表位等。对这些表位的分子组成及结构、在体液免疫应答反应中表现的研究，可以进一步提高我们对表位生物学意义的认识，加深对病毒感染过程中抗原性蛋白功能的理解；并为最终解决致病性病毒提供有用的工具和手段

在**EIAV**病毒中，**Gp90**是个高度糖基化的蛋白质，是诱导宿主产生中和抗体的主要抗原，因此我们想对这个糖蛋白进行表位作图，先用生物信息学技术进行下抗原表位的预测，为指导后续实验的设计和进行

1. 抗原表位预测 (BcePred)

BcePred可以预测连续B细胞表位，采用的理化参数包括亲水性、柔韧性、可及性、极性、暴露表面、转角和抗原性



Home	submission	Help	Output format	Algoritha	Team
----------------------	----------------------------	----------------------	-------------------------------	----------------------------------	----------------------

SUBMISSION FORM

Sequence name (optional) :

Paste your sequence below:

(Amino acid sequence in one lettercode. No header line)

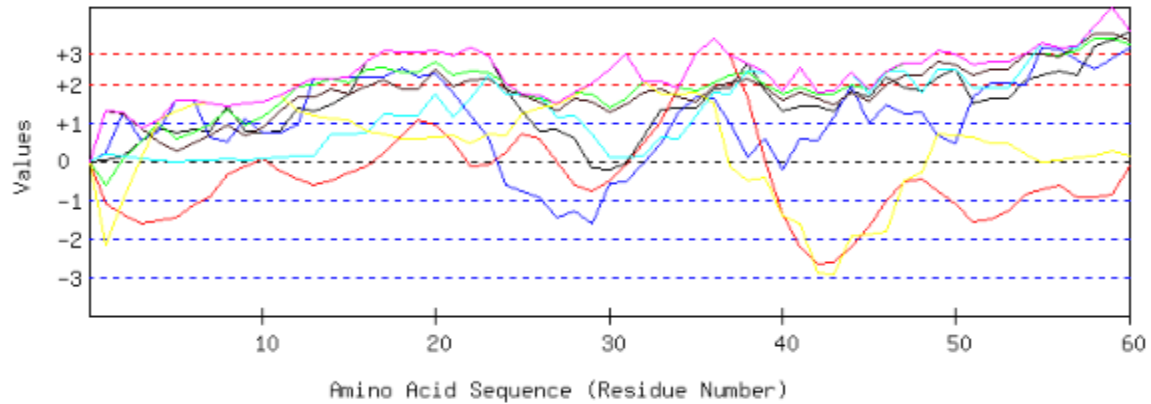
Or Subait sequences from file :

Threshold [-3 to 3] :

Hydrophilicity:
Accessibility:
Exposed Surface:
Antegenic Propensity:

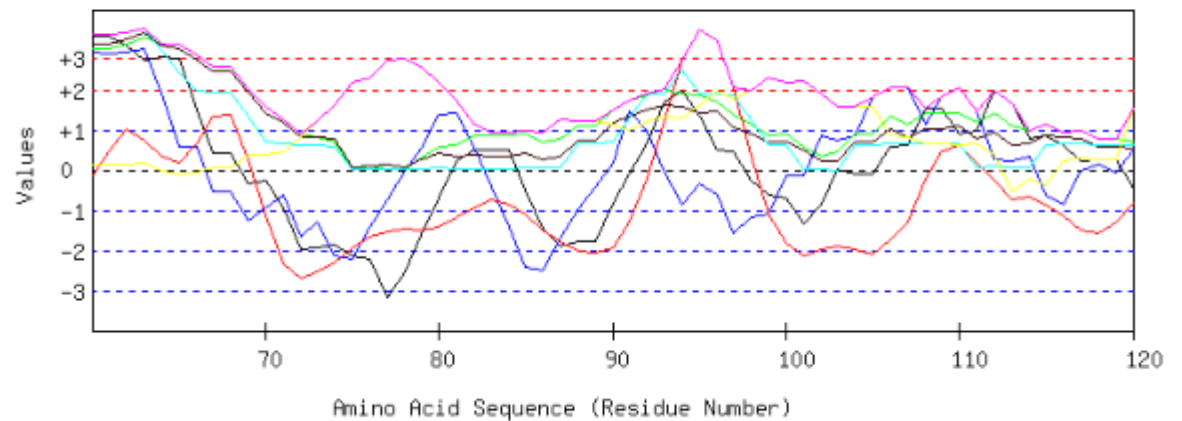
Flexibility:
Turns:
Polarity:
Combined:

GRAPHICAL RESULT :: SEQ 1 to 60



Hydrophilicity Turns Surface Flexibility Polar Accessibility Antigenic Comb4 Combined

GRAPHICAL RESULT :: SEQ 61 to 120



Hydrophilicity Turns Surface Flexibility Polar Accessibility Antigenic Comb4 Combined

2. 抗原表位预测 (ABCpred)

ABCpred服务器是在机器学习的基础上开发连续性B细胞表位预测的工具，其训练集来自Bcipep数据库中的B细胞表位



Artificial neural network based B-cell epitope prediction server

[Home](#)

[Submission](#)

[Help](#)

[Method](#)

[Team](#)

SUBMISSION FORM

Sequence name (optional) :

Paste your sequence below:

(Amino acid sequence in one lettercode. No header line)

Or Submit sequences from file :

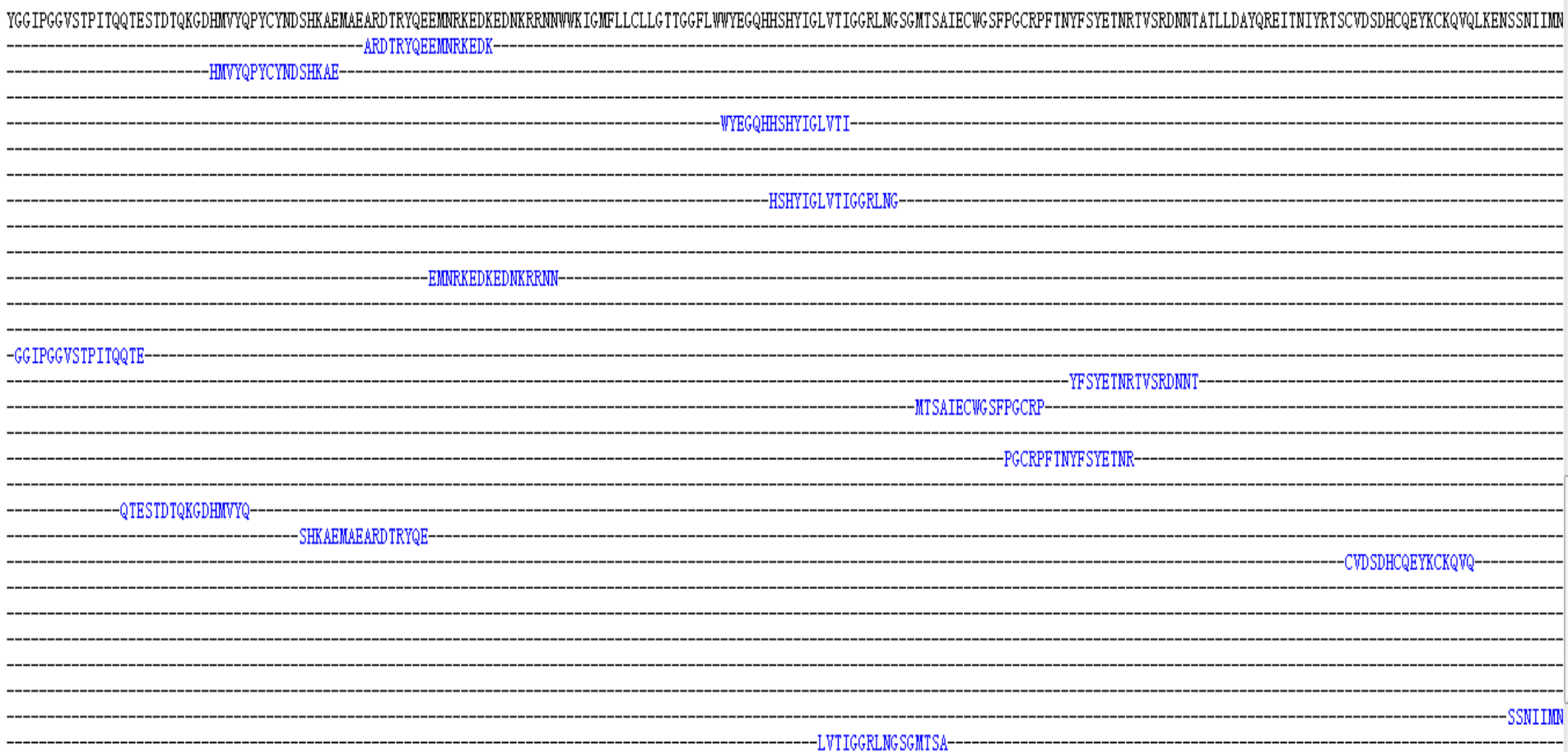
Threshold [0.1 to 1] :

Select a window length to use for prediction:

Overlapping filter: ON OFF

Rank	Sequence	Start position	Score
1	ARDTRYQEEMNRKEDK	45	0.91
1	HMVYQPYCYND SHKAE	26	0.91
2	MQESDNNTWIPKRCNE	226	0.90
3	WYEGQHSHYIGLVTI	89	0.89
3	RTNIKRDPDYSGFYQVP	311	0.89
3	IRMCVQPPFFLFKQND	260	0.89
4	HSYIGLVTIGGRLNG	95	0.88
5	KTSGITPLPISSTANT	422	0.87
5	TELQWQLVKTSGITP	413	0.87

OVERLAP DISPLAY



选取得分>0.85以上的表位片段

Rank	Sequence	Start position	Score
1	ARDTRYQEEMNRKEDK	45	0.91
1	HMVYQPYCYNDSHKAE	26	0.91
2	MQESDNNTWIPKRCNE	226	0.90
3	WYEGQHSHYIGLVTI	89	0.89
3	RTNIKRPDYSGFYQVP	311	0.89
3	IRMCVQPPFFLFKQND	260	0.89
4	HSHYIGLVTIGGRLNG	95	0.88
5	KTSGITPLPISSTANT	422	0.87
5	TELQWQLVKITSGITP	413	0.87
6	EMNRKEDKEDNKRRNN	53	0.86
7	IIMYESNNVQYLLCNT	347	0.85
7	LVPDIEMQESDNNTWI	220	0.85
7	GGIPGGVSTPITQQTE	2	0.85
7	YFSYETNRTVSRDNNT	132	0.85
7	MTSAIECWGSFPGCRP	113	0.85

总 结

通过gp90蛋白进行生物信息学分析，对这个蛋白有了基本的了解，同时预测出可能的抗原表位，为后续抗原表位鉴定试验提供了根本性的指导!

Reference

1. Howe L, Leroux C, Issel C, et al. Equine infectious anemia virus envelope evolution in vivo during persistent infection progressively increases resistance to in vitro serum antibody neutralization as a dominant phenotype [J]. *J Virol*, 2002, 76: 10588-10597.
2. Johnson W E, Sanford H, Schwall L, et al. Assorted mutations in the envelope gene of simian immunodeficiency virus lead to loss of neutralization resistance against antibodies representing a broad spectrum of specificities [J]. *J Virol*, 2003, 77(18): 9993-10003.
3. Ball J M, Swaggerty C L, Pei X, et al. SU proteins from virulent and avirulent EIAV demonstrate distinct biological properties. *Virology* [J]. 2005, 333(1): 132-144.
4. Bolmstedt A, Sjolander S, Hansen J E, et al. Influence of N-linked glycans in V4-V5 region of human immunodeficiency virus type 1 glycoprotein gp160 on induction of a virus-neutralizing humoral response [J]. *J Acquir Immune Defic Syndr Hum Retrovirol*, 1996, 12(3): 213-220.

致谢

感谢罗老师！

感谢小组各位成员！

感谢农科院！

感谢北京大学！

Half day on the Web, saves you half month in the lab!