

樱桃谷肉鸭 α -干扰素（IFN- α ）生物信息学分析

G08:

岳嘉蘋 王艳歌

杨洋 张浩浩

2012年6月10号

- 背景知识
- 编码区核酸序列分析

密码子使用
读码框结构
限制位点
核酸进化树

- 蛋白序列分析

理化性质
二级结构
蛋白序列进化树
同源建模

背景知识

干扰素（interferon, IFN）是一组具有多功能活性的细胞因子，具有抗病毒、抗肿瘤和免疫调节等作用。其种类主要包括IFN- α ，IFN- β 和IFN- γ ，许多干扰素制剂已被广泛的应用到临床上。

背景知识

近年来禽干扰素的研究逐渐受到重视，目前为止在禽类中只发现了IFN- α 和IFN- γ 。鸭的病毒性疾病如鸭瘟、病毒性肝炎等，对养鸭业危害严重，干扰素对这些疾病有一定的治疗作用，这也是对鸭干扰素研究的意义所在。

[back](#)

核酸序列分析

利用RT-PCR方法，克隆出的樱桃谷肉鸭IFN- α 的基因编码区测序结果为

编码区序列长
576bp

```
>D-alpha+  
TTAGCGCATGGTGCGGGTGAGGCGGTGGATGCGCTGGAAGCAGGCGTGAGCCTCG+  
AGGCGGACGTGGTCCCATGCGCAGGGGCTGTAGGTGTGGTTCTGGAGGAAGTGTT+  
GGATGCAGCCGAAGTACTTGTGATGCTGAGGTGAAGGTTGCGGGGCCCTCGCCT+  
GTGGAGGCGCGCGGCCGTCGGCTGGGAAGCAGCGCTCGAGGTGGTGGATGTGGTGC+  
TGAAGCTGGTTGAGGAGGTCGTGGCGTGCGGTGTGGAGCCAGTGCGCGGGGGTGC+  
TGGGGCTGCTGAGGGTGTGCGAAGAGGTGTTGGAGGAGGTGGAGGGCGGTGTGTGC+  
GGCTTGCTGCGTGTGCGTTGGTGTCCAGGGGGGTGTCCGGGAAGGAGCAAGGCGCG+  
TGTTGCTGCGGGCAGGGCTGTGTGGGGCTGGGAGCCATGTTGCGGAGGAGCTGGA+  
GGCTCTCCAGGCGAAGGCGCTGTGCGTGAGGCGCAGGGGGCTGCAGGAGAAGGC+  
GTTGGCGGGAGGCGTGAGGAGGAGCAGGAGGGCCAGGGCGCTGTAGATGGCTGGT+  
GGTGGTGGGGCTGATGGCCCAGGCAT+
```

核酸序列分析

---密码子使用

- 用WebLab中[cusp\(v6.0.1\)](#) 软件分析樱桃谷肉鸭IFN- α 编码区的密码子使用:

```
#Coding GC 67.53%  
#1st letter GC 87.50%  
#2nd letter GC 48.44%  
#3rd letter GC 66.67%
```

#Codon	AA	Fraction	Frequency	Number
GCA	A	0.241	36.458	7
GCC	A	0.034	5.208	1
GCG	A	0.379	57.292	11
GCT	A	0.345	52.083	10
TGC	C	0.750	15.625	3
TGT	C	0.250	5.208	1
GAC	D	0.167	5.208	1
GAT	D	0.833	26.042	5
GAA	E	0.267	41.667	8
GAG	E	0.733	114.583	22
TTC	F	0.000	0.000	0
TTT	F	0.000	0.000	0
GGA	G	0.059	10.417	2
GGC	G	0.382	67.708	13
GGG	G	0.294	52.083	10
GGT	G	0.265	46.875	9
CAC	H	0.000	0.000	0
CAT	H	1.000	15.625	3
ATA	I	0.000	0.000	0
ATC	I	0.000	0.000	0
ATT	I	0.000	0.000	0
AAA	K	0.000	0.000	0
AAG	K	1.000	10.417	2
CTA	L	0.000	0.000	0
CTC	L	0.176	15.625	3
CTG	L	0.471	41.667	8

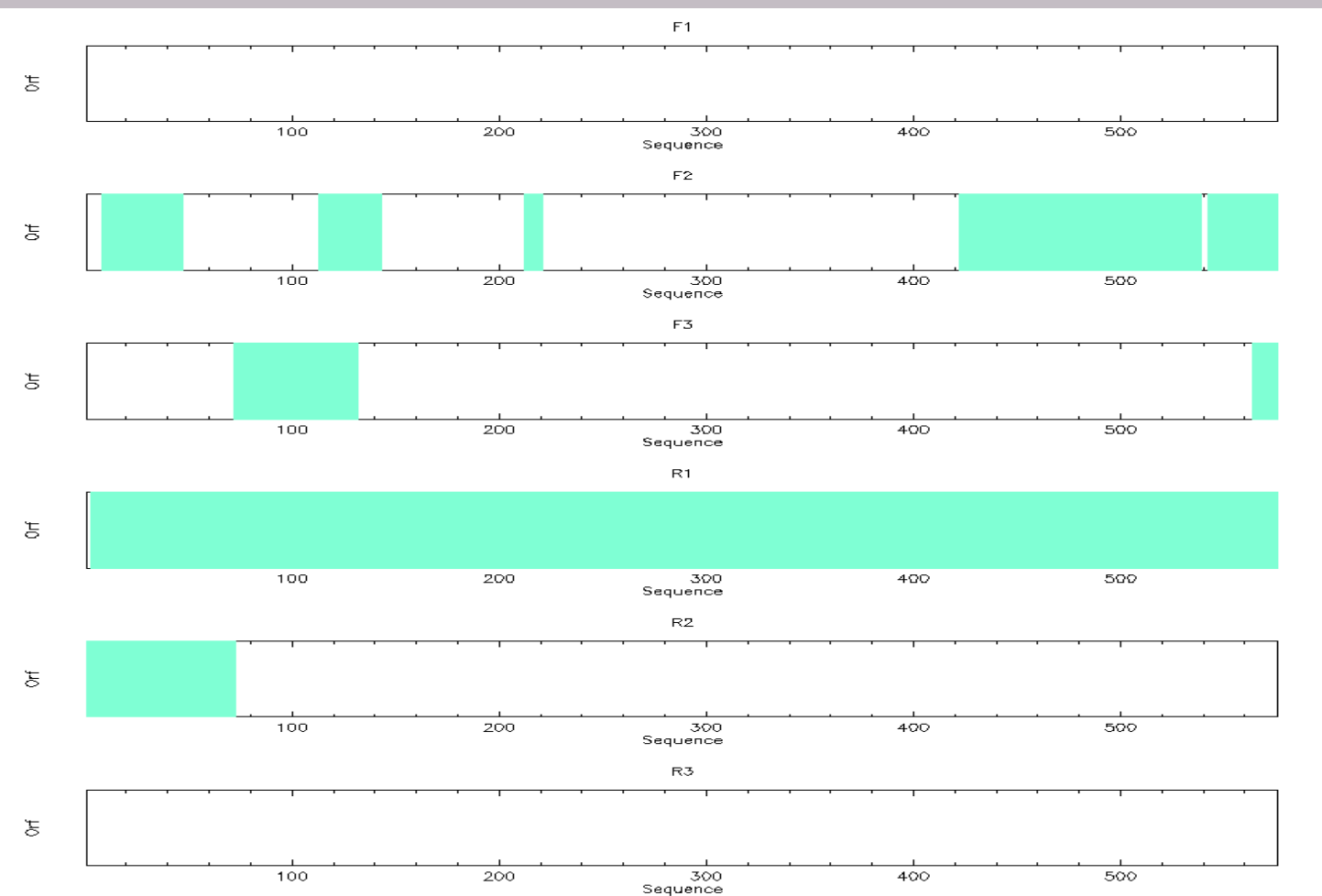
CTT	L	0.059	5.208	1
TTA	L	0.059	5.208	1
TTG	L	0.235	20.833	4
ATG	M	0.000	0.000	0
AAC	N	0.000	0.000	0
AAT	N	0.000	0.000	0
CCA	P	0.500	15.625	3
CCC	P	0.333	10.417	2
CCG	P	0.000	0.000	0
CCT	P	0.167	5.208	1
CAA	Q	0.000	0.000	0
CAG	Q	1.000	20.833	4
AGA	R	0.000	0.000	0
AGG	R	0.300	15.625	3
CGA	R	0.000	0.000	0
CGC	R	0.300	15.625	3
CGG	R	0.200	10.417	2
CGT	R	0.200	10.417	2
AGC	S	0.667	10.417	2
AGT	S	0.000	0.000	0
TCA	S	0.000	0.000	0
TCC	S	0.000	0.000	0
TCG	S	0.333	5.208	1
TCT	S	0.000	0.000	0
ACA	T	0.000	0.000	0
ACC	T	0.000	0.000	0
ACG	T	0.000	0.000	0
ACT	T	0.000	0.000	0
GTA	V	0.081	15.625	3
GTC	V	0.216	41.667	8

GTG	V	0.514	98.958	19
GTT	V	0.189	36.458	7
TGG	W	1.000	31.250	6
TAC	Y	0.000	0.000	0
TAT	Y	0.000	0.000	0
TAA	*	0.000	0.000	0
TAG	*	0.000	0.000	0
TGA	*	1.000	5.208	1

核酸序列分析

---读码框

- 用WebLab 中[plotorf\(v6.0.1\)](#)软件预测该序列可能的读码框，只有反义链读码结构完整



六种可能的读码框中只有第一种反义链读码框完整。

核酸序列分析

- 用[showorf\(v6.0.1\)](#) 软件分析该反义链的序列特征

SHOWORF of D-alpha from 1 to 576

```
-----|-----|-----|-----|-----|-----|-----|-----|
1 TTAGCGCATGGTGCGGGTGAGGCGGTGGATGCGCTGGAAGCAGGCGTGAGCCTCGAGGCGGACGTGGTCCCATGCGCAGGGGCTGTAGGT 90
R1 191 * R M T R T L R H I R Q F C A H A E L R V H D W A C P S Y T 163

-----|-----|-----|-----|-----|-----|-----|-----|
91 GTGGTTCTGGAGGAAGTGTGGATGCAGCCGAAGTACTTGTGTGATGCTGAGGTGAAGGTTGCGGGGCCCTCGCCTGTGGAGGCGCGCGGC 180
R1 162 H N Q L F H Q I C G F Y K N I S L H L N R P G R R H L R A A 133

-----|-----|-----|-----|-----|-----|-----|-----|
181 GTCGGCTGGGAAGCAGCGCTCGAGGTGGTGGATGTGGTGCTGAAGCTGGTTGAGGAGGTCGTGGCGTGGGTGTGGAGCCAGTGC GCGGG 270
R1 132 D A P F C R E L H H I H H Q L Q N L L D H R A T H L W H A P 103

-----|-----|-----|-----|-----|-----|-----|-----|
271 GGTGCTGGGGCTGCTGAGGGTGTGCAAGAGGTGTTGGAGGAGGTGGAGGGCGGTGTGTGCGGCTTGTGCGTGTGCGTGGTGTCCAGGGG 360
R1 102 T S P S S L T D F L H Q L L H L A T H A A Q Q T D N T D L P 73

-----|-----|-----|-----|-----|-----|-----|-----|
361 GGTGTCCGGGAAGGAGCAAGGCGCGTGTGCTGCGGGCAGGGCTGTGTGGGGCTGGGAGCCATGTTGCGGAGGAGCTGGAGGCTCTCCA 450
R1 72 T D P F S C P A H Q Q P C P Q T P S P A M N R L L Q L S E W 43

-----|-----|-----|-----|-----|-----|-----|-----|
451 GGCGAAGGCGCTGTGCTGGAGGCGCAGGGGCTGCAGGAGAAGGCGTTGGCGGGAGGCGTGAGGAGGAGCAGGAGGGCCAGGGCGCTGTA 540
R1 42 A F A S D H L R L P S C S F A N A P P T L L L L L A L A S Y 13

-----|-----|-----|-----|
541 GATGGCTGGTGGTGGTGGGGCTGATGGCCCAGGCAT 576
R1 12 I A P P P P A S P G P M 1
```

可以看出，
其编码区除
终止密码子
外共编码了
191个氨基
酸。

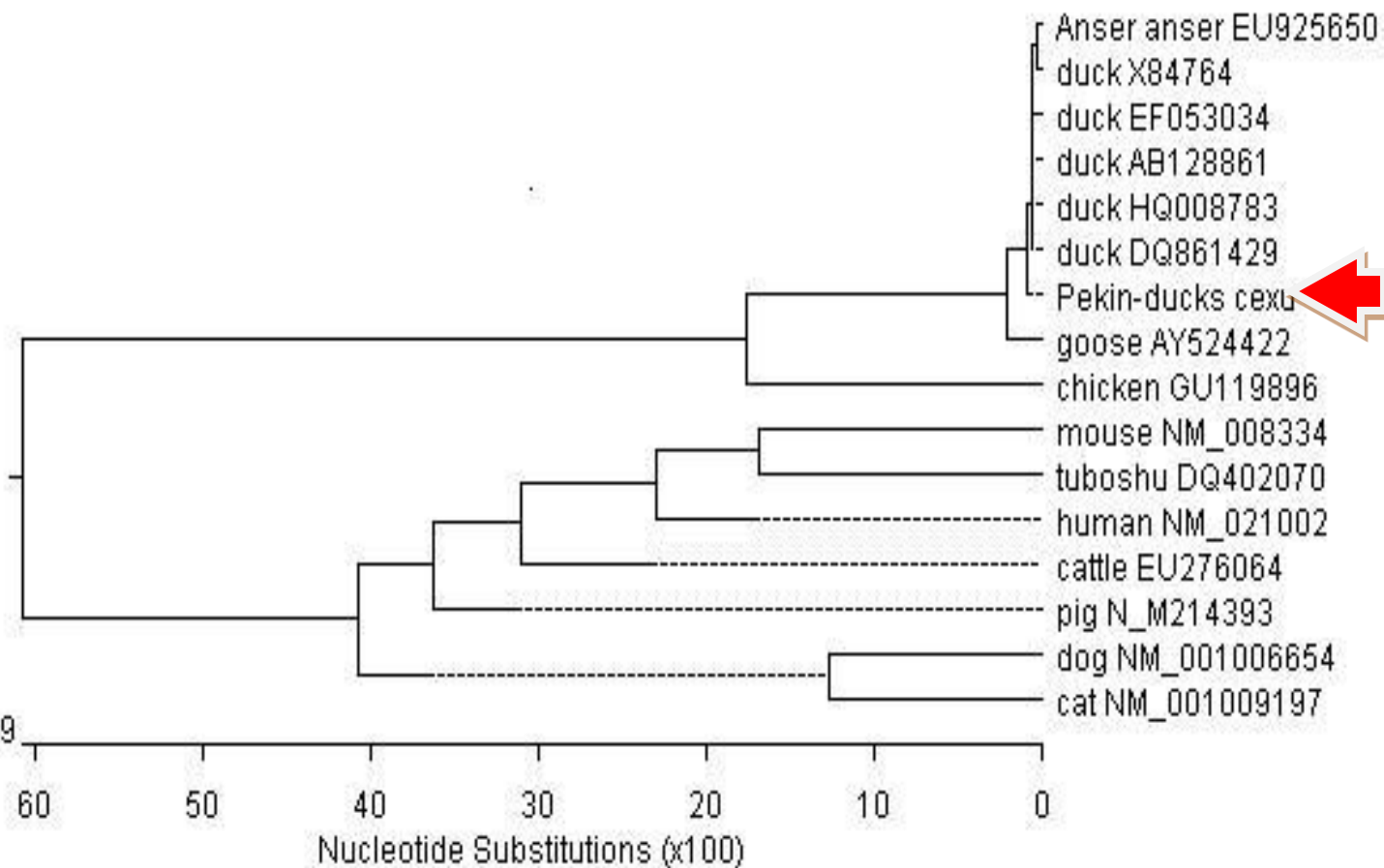
核酸序列分析

---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1021个、226个、67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。

核酸序列分析

---核酸序列进化树



大体上分为两大枝，一枝为禽类，另一枝为哺乳类动物，樱桃谷肉鸭与鹅、雁的亲缘关系较近，与鸡的亲缘关系相对次之。

back

蛋白序列分析

---理化性质

利用WebLab中[pepstats\(v6.0.1\)](#)和ExPASy网站ProtParam软件
分析该干扰素的理化性质

氨基酸数目: 191

分子式为 $C_{965}H_{1475}N_{291}O_{261}S_{10}$

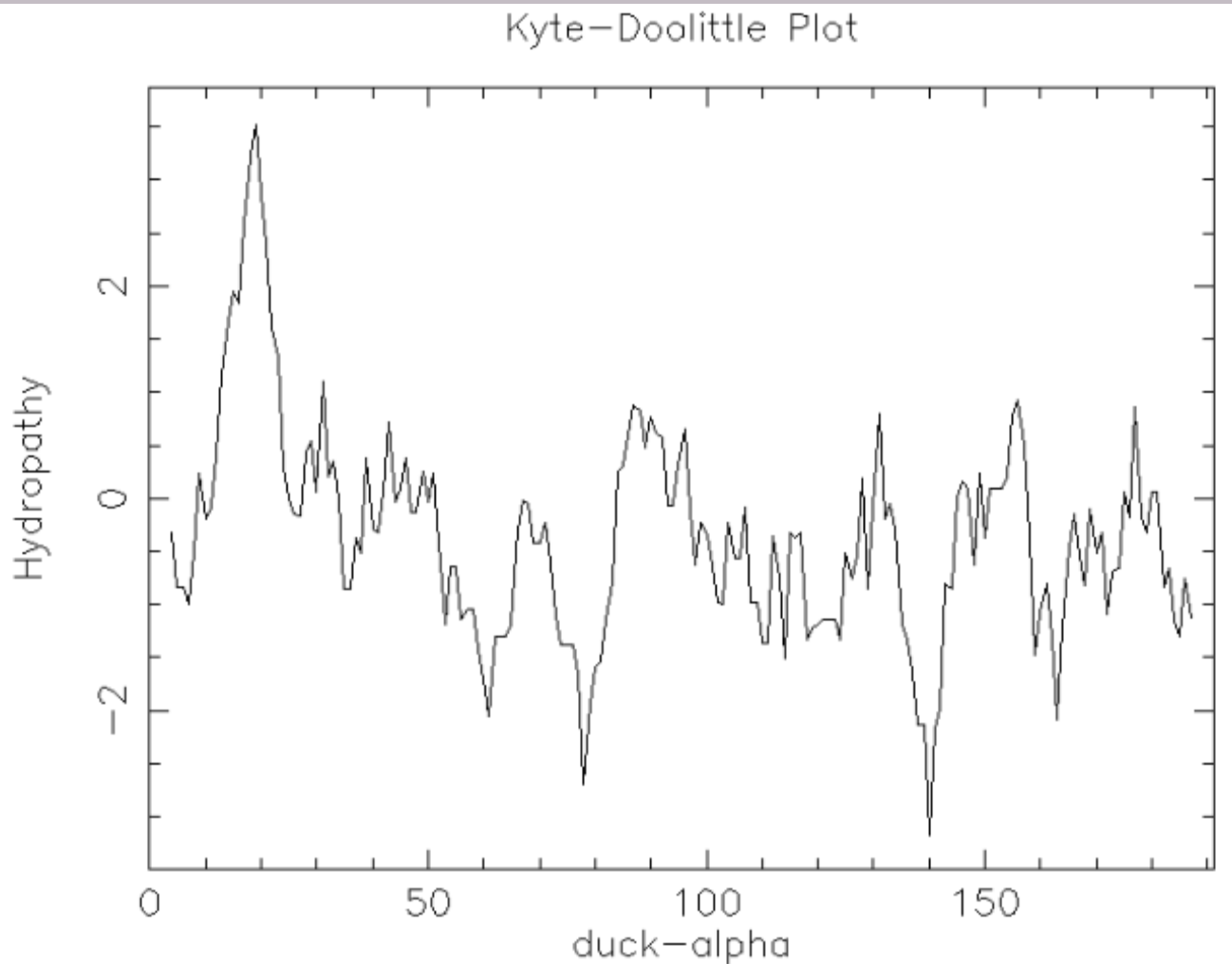
分子量: 21, 649.7Da

等电点: 8.58

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	56	29.319
Small	(A+B+C+D+G+N+P+S+T+V)	93	48.691
Aliphatic	(A+I+L+V)	55	28.796
Aromatic	(F+H+W+Y)	33	17.277
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	103	53.927
Polar	(D+E+H+K+N+Q+R+S+T+Z)	88	46.073
Charged	(B+D+E+H+K+R+Z)	44	23.037
Basic	(H+K+R)	33	17.277
Acidic	(B+D+E+Z)	11	5.759

蛋白质序列分析

- 利用[pepwindow\(v6.0.1\)](#) 软件对干扰素进行疏水性亲水性分析

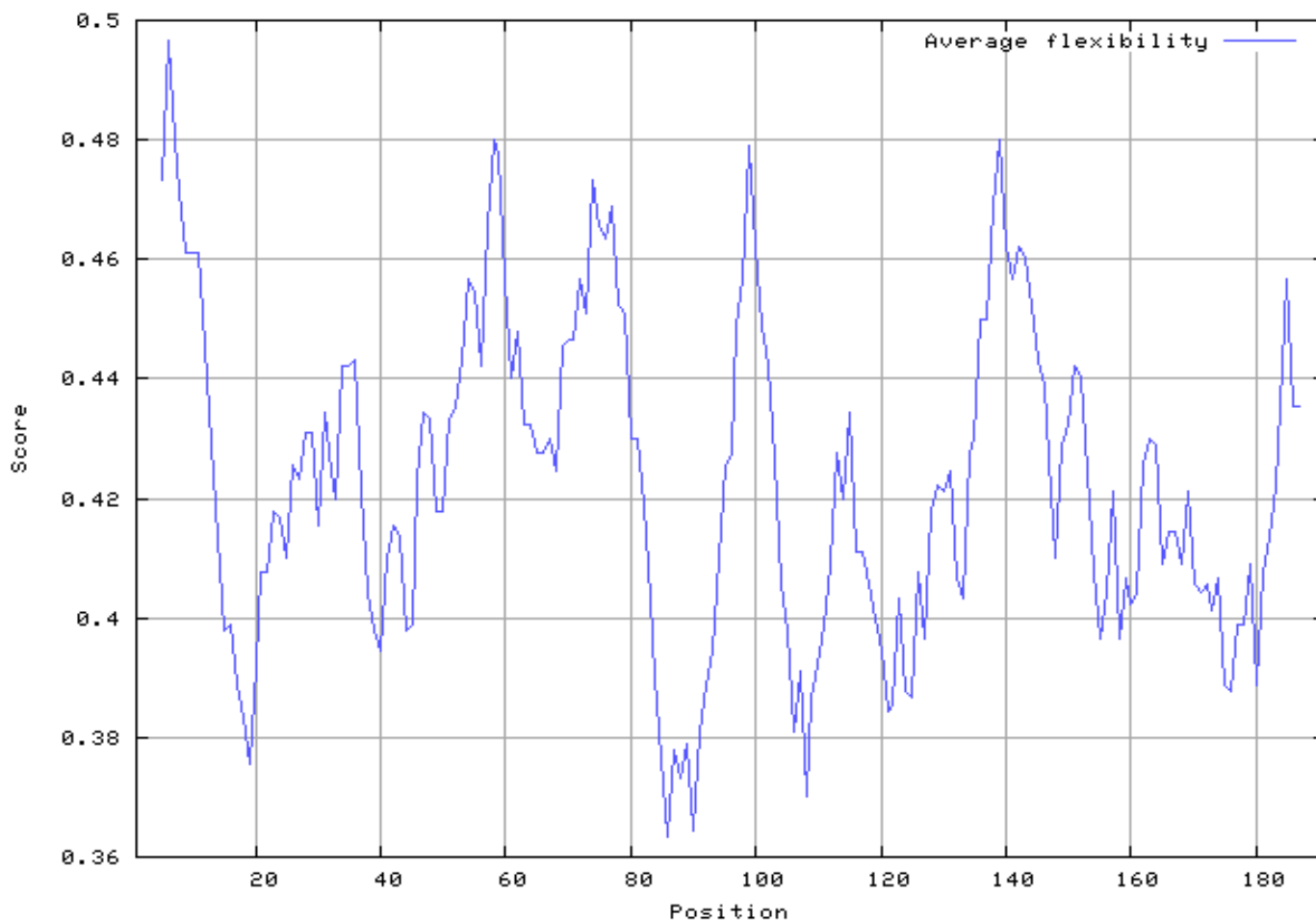


大部分序列的亲水性得分较低，只有第**20**氨基酸附近的序列具有亲水性。

蛋白质序列分析

利用ProtScale对其进行柔性刚性分析

ProtScale output for user sequence



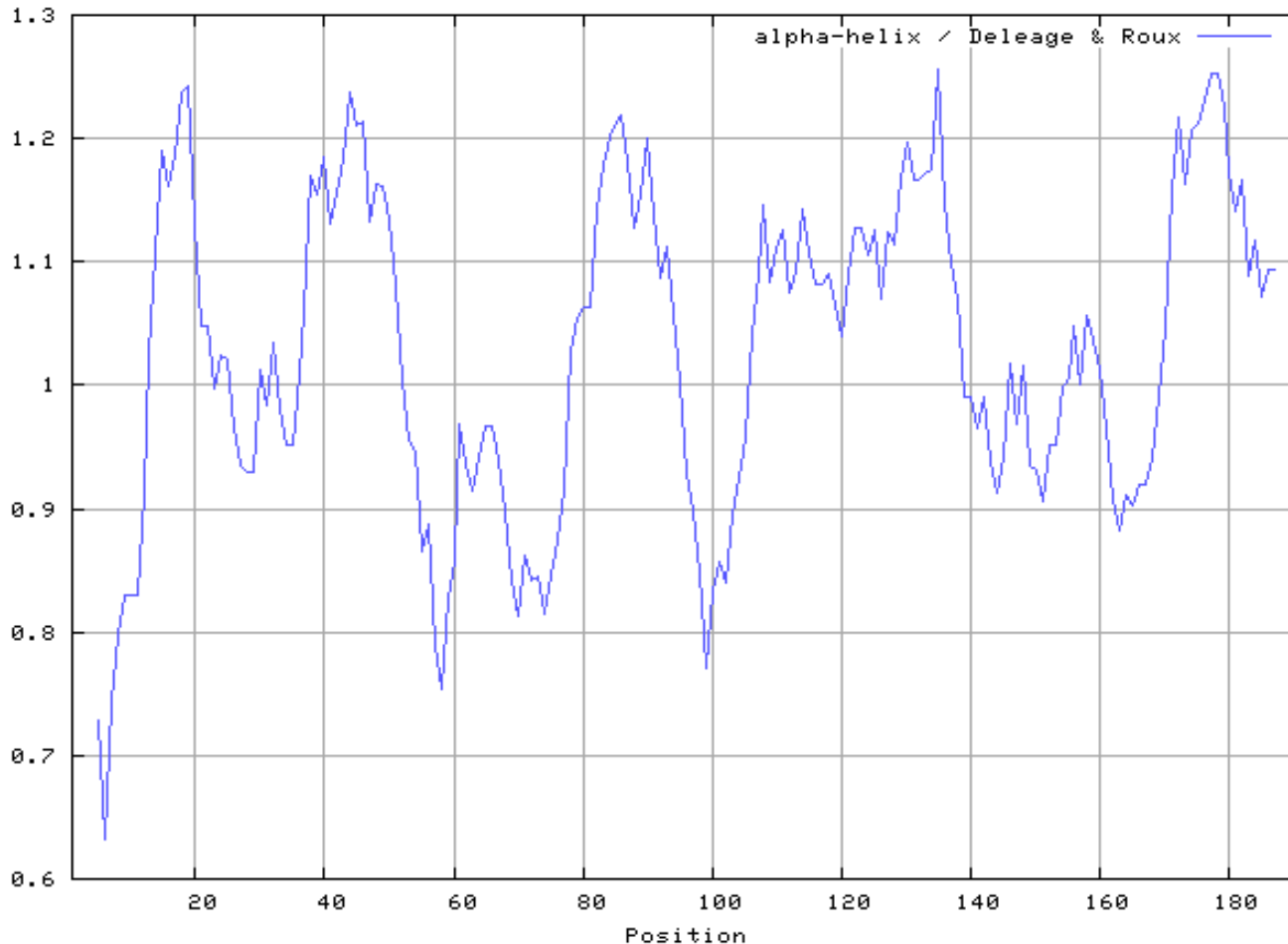
有**5**个柔性较大的区域，可能是高级结构中的不规则卷曲部分。

蛋白质序列分析

---二级结构

利用ProtScale软件进行 α 螺旋预测

ProtScale output for user sequence

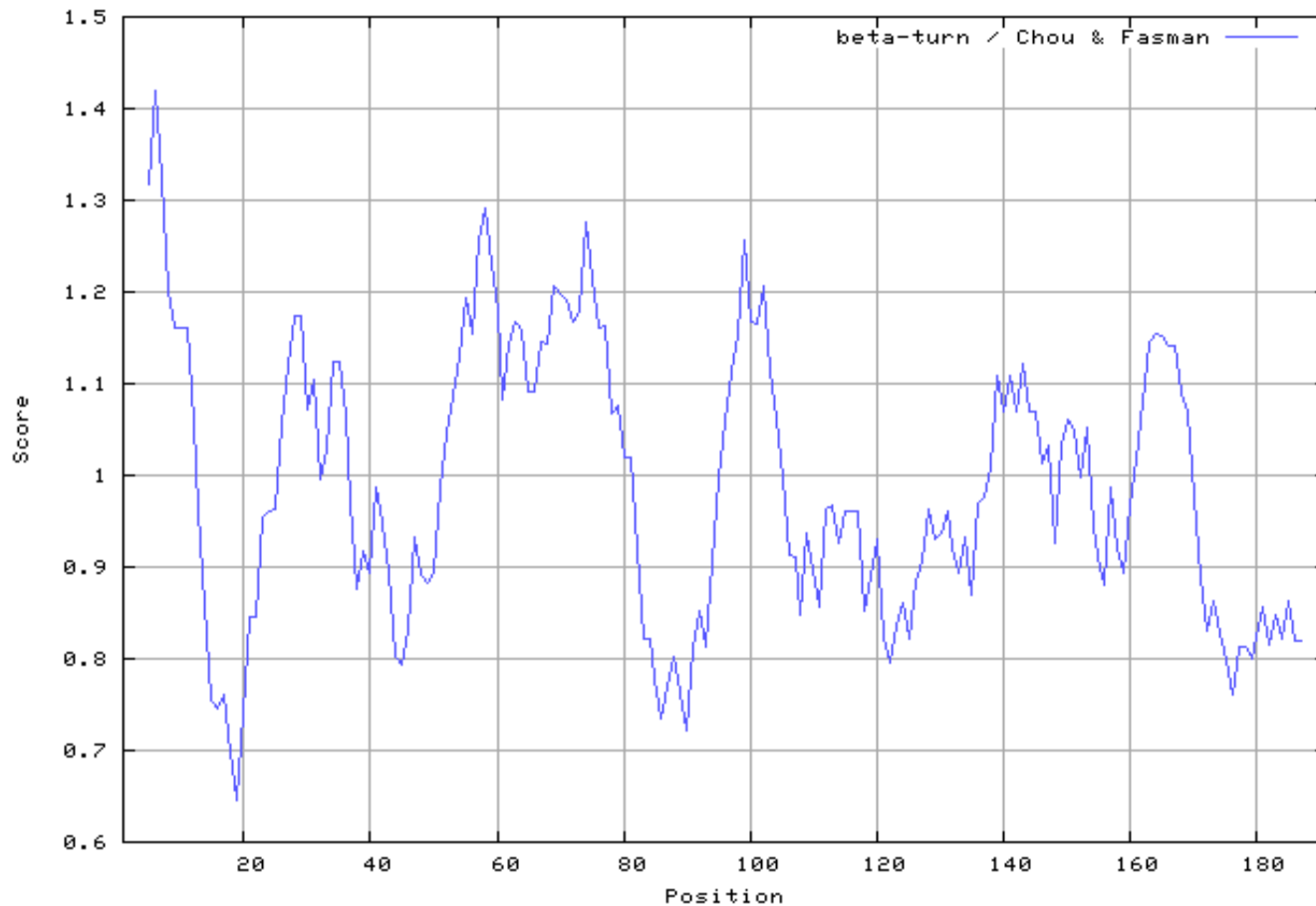


有5个明显的 α 螺旋区域，这与同源建模结果一致。

蛋白质序列分析

β 转角预测

ProtScale output for user sequence

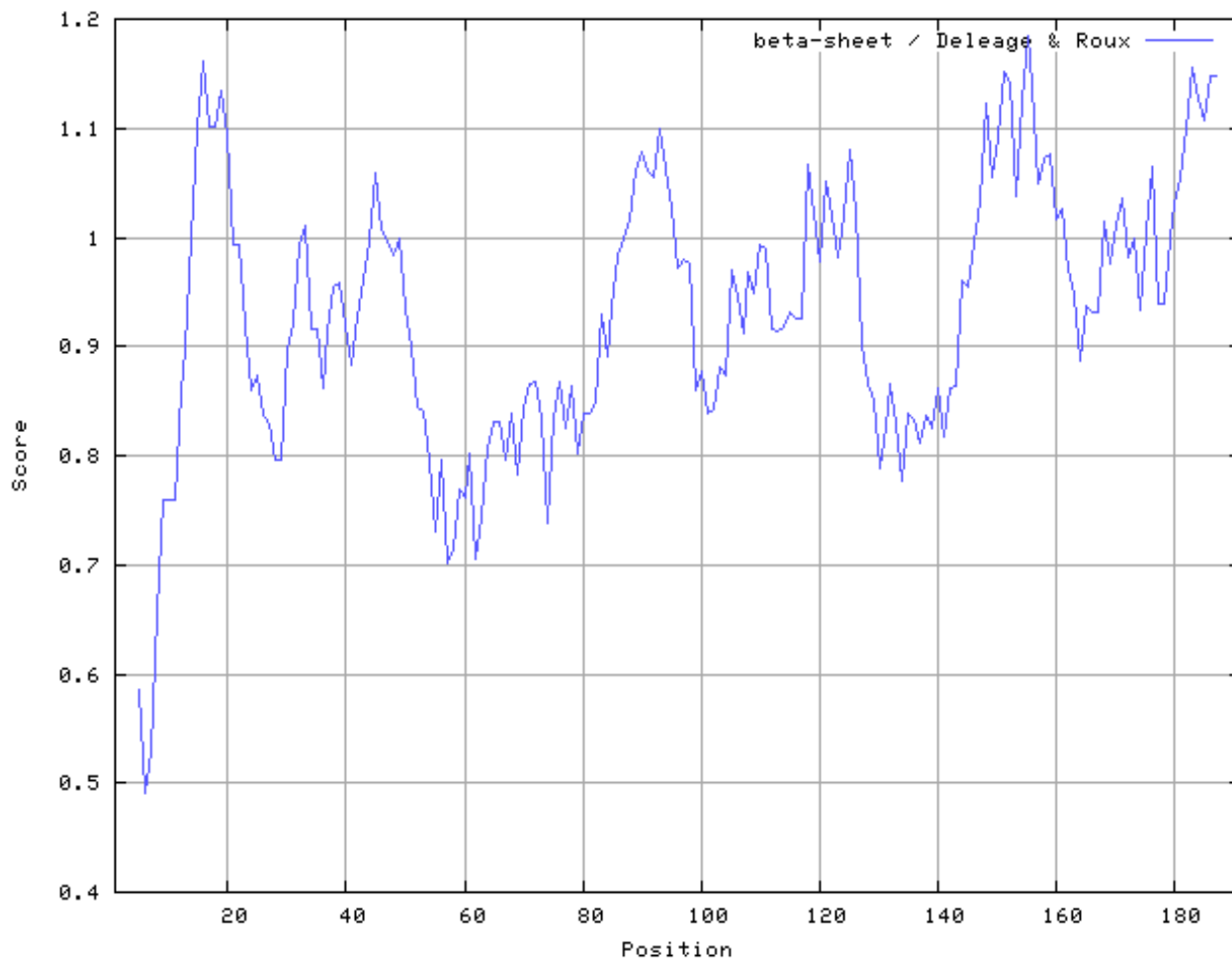


有四个可能的 β 转角区域。

蛋白质序列分析

β 折叠预测

ProtScale output for user sequence



存在三个可能的可能的 β 折叠区域，但后面同源建模预测的结果中并不存在 β 折叠，这也提示我们生物信息学软件只是用来进行预测，我们需要结合多种软件进行综合分析，而不能偏信某一个软件的分析结果。

利用WebLab中garnier软件进行二级结构预测，与ExPASy网站结果大致相同。



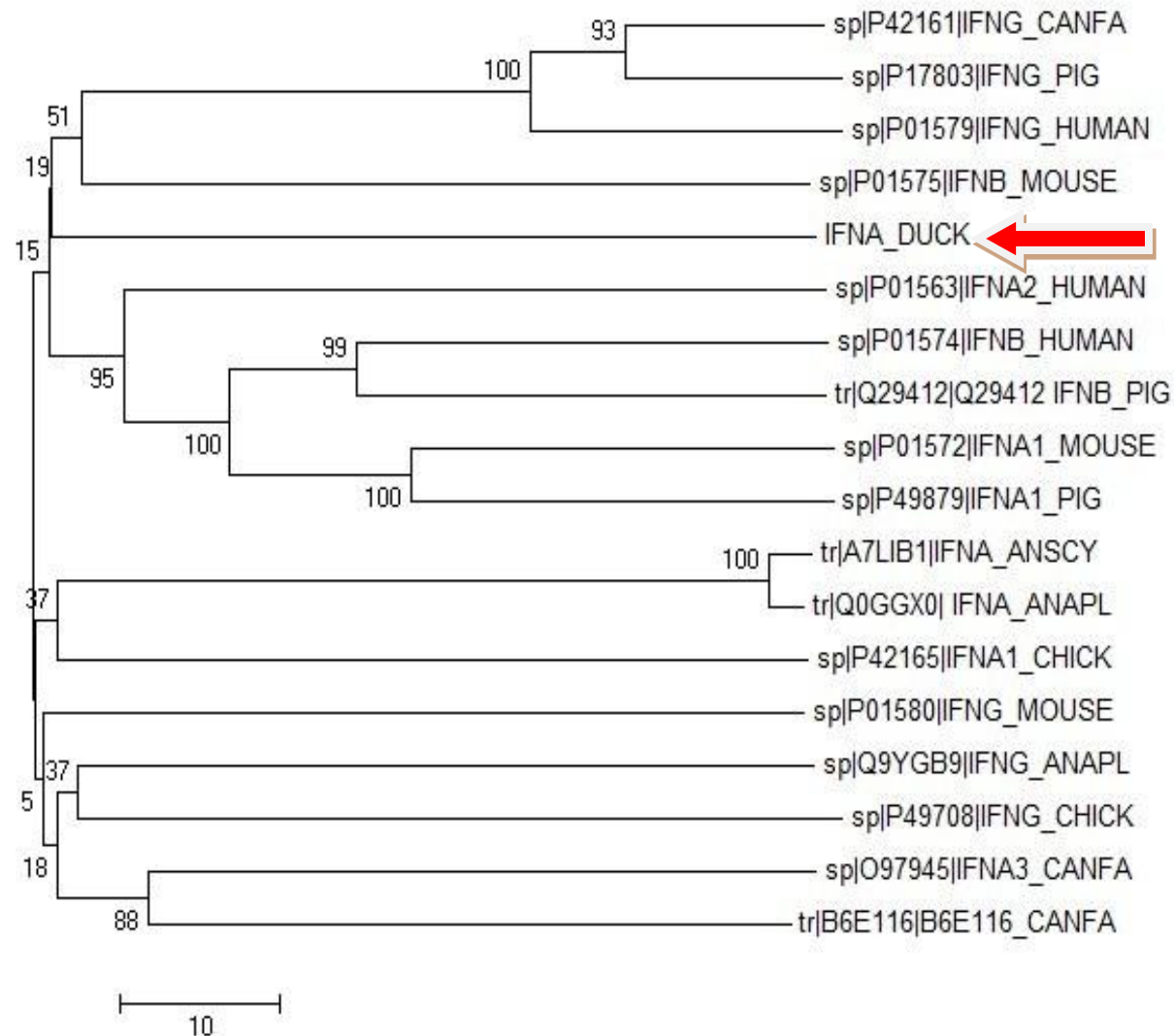
```

      .   10   .   20   .   30   .   40   .   50
MFGPSAPPFPAIYSALALLLLTPPANAFSCSPLRLHDSAFAWESLQLLR
helix          HHHH          HH   HHHHHHHHH
sheet          EEEE   EEEE          E
turns         T          TTTTTT TT          T
coil  CCC  CCCCCC          CC          CCC          C
      .   60   .   70   .   80   .   90   .  100
NMAPSPTQPCPQQHAPCSFPDTPLDTNDTQQAHTALHLLQHLFDTLSSP
helix                                HHHHHHHHHHHHH
sheet                                EE
turns  T  T TT TTTTTT TTT TT  TTT          TTT  T
coil  CC C  C          C  C  CC  CC          CC
      .  110   .  120   .  130   .  140   .  150
STPAHWLHTARHDLLNQLQHIIHHLERCFFPADAARLHRRGPRNLHLSINK
helix          HHHHHH   HHHHHHHH   HHHHHHHH
sheet                                EEEE
turns          TTT          TTTTTT          TTT TT          TTT
coil  CCCC          C          C  C
      .  160   .  170   .  180   .  190
YFGCIQHFLQNHTYSPCAWDHVRLEAHACFPQRIHRLTRTMR
helix                                HHHHHHHHHHHHH
sheet          EEEEE
turns  TTT          TTTTTT TTT          TT TTTTT
coil          C          C

```

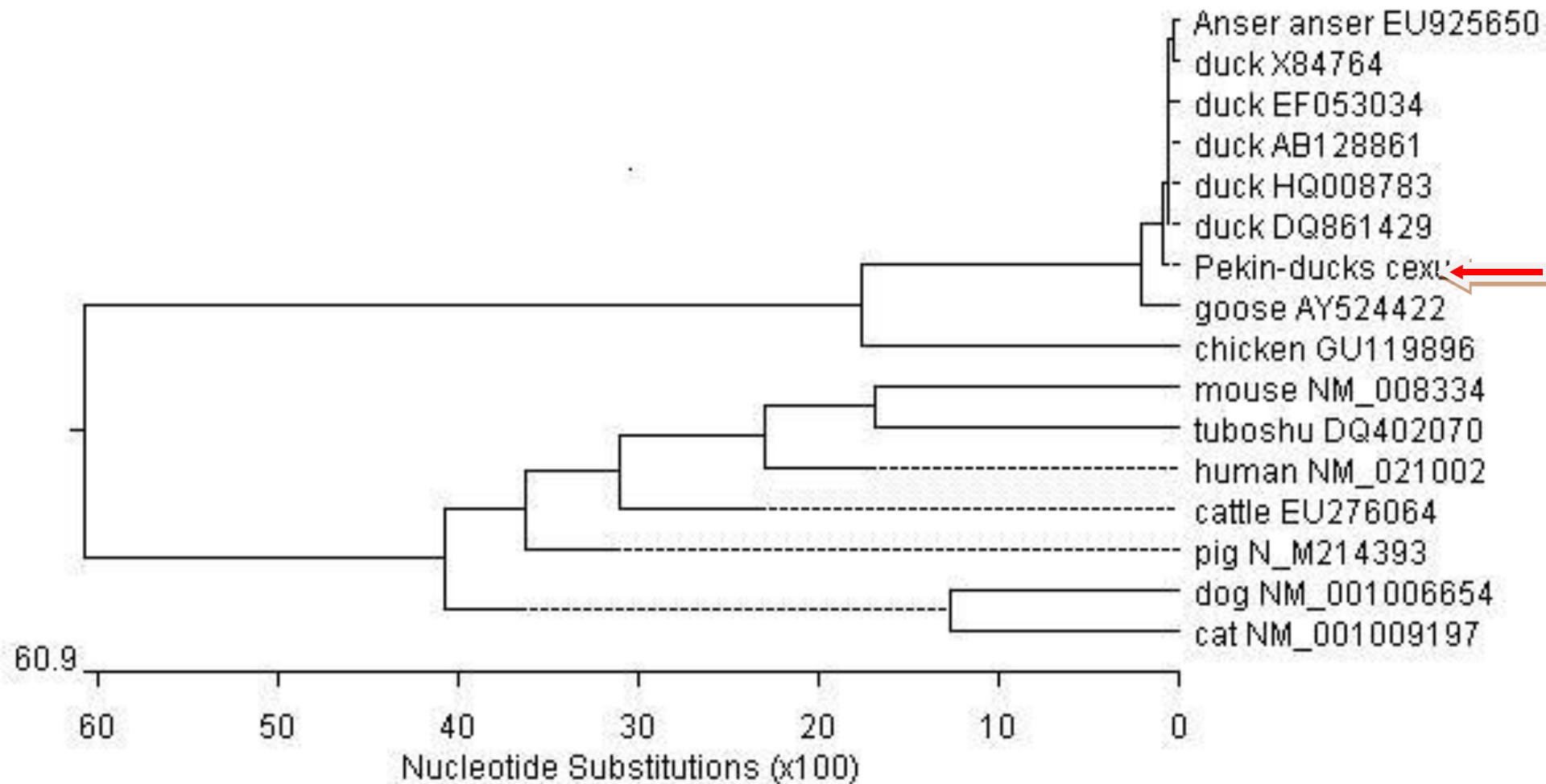
蛋白序列分析

---进化树



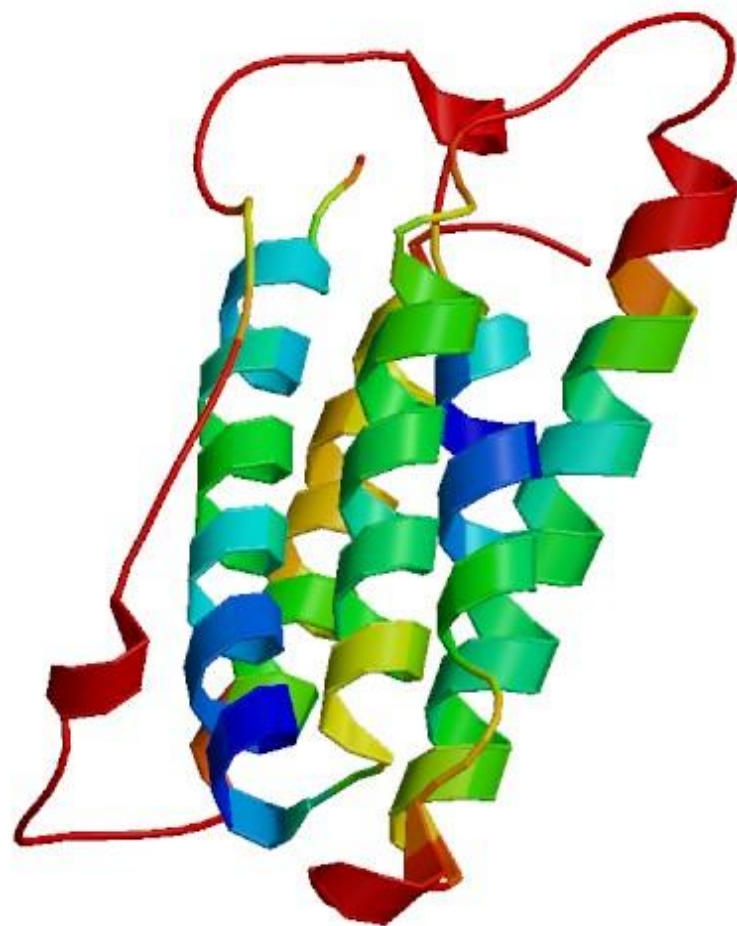
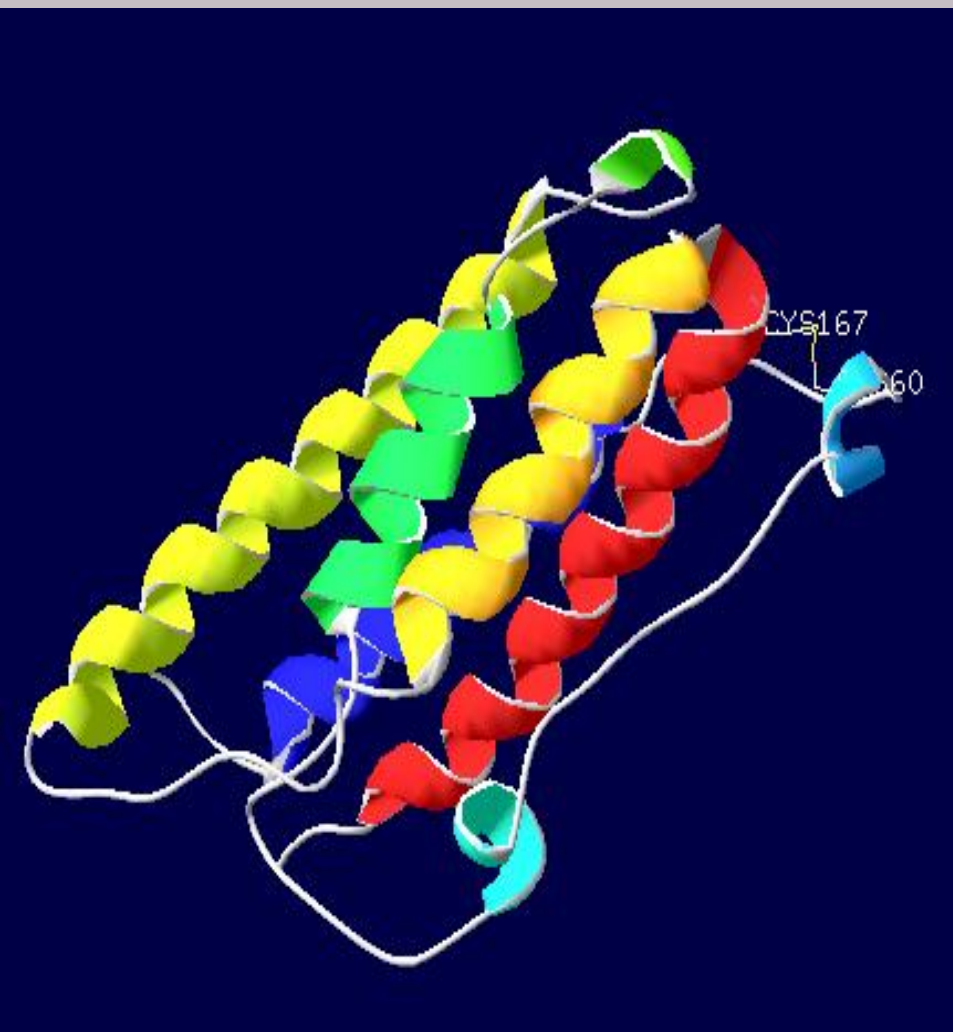
在蛋白水平上樱桃谷肉鸭的IFN- α 与鼠、人等动物的亲缘关系较近，而与鸡、雁的亲缘关系较远。这说明由于密码子的简并性，使得核酸序列分析结果不同于氨基酸序列的分析结果。

---核酸序列进化树



蛋白序列分析

利用swiss-model网站进行建模



致谢！

- 在此代表我们组的组员，感谢罗老师在这个学期对我们的悉心的教导，你使收获良多！
- 同时，谢谢全体同学，在这个学期中，跟你们的交流让我们也学到了很多很多。谢谢！



*Thanks for your
attention !!!*