

猪碳酸酐酶3生物信息学相关分析

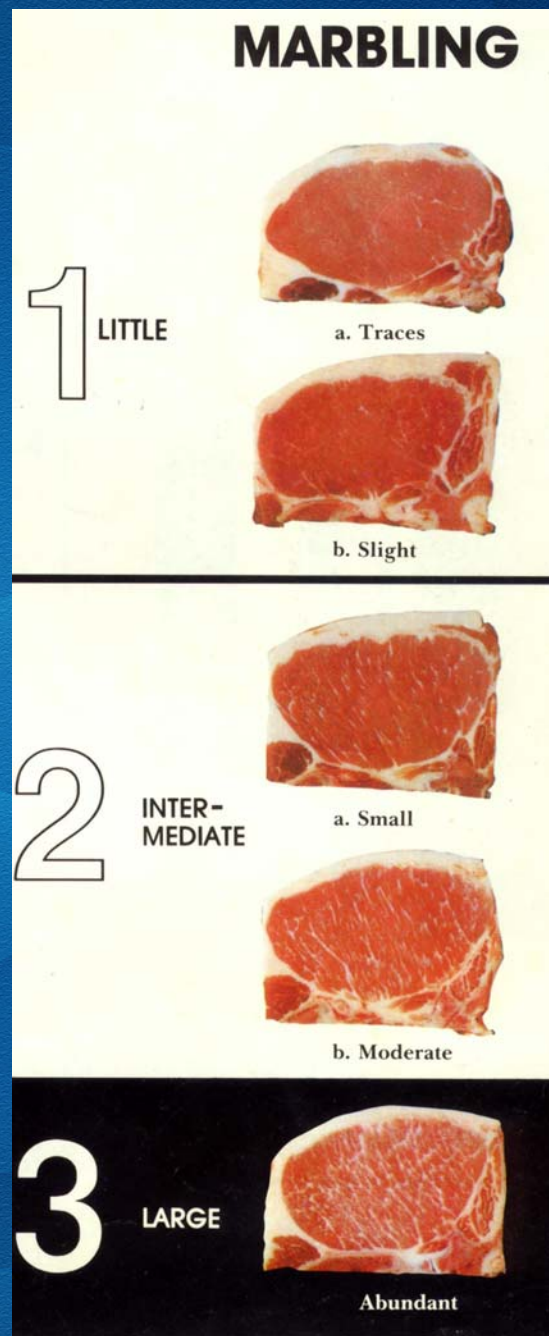
何晓红

中国农科院北京畜牧兽医研究所

2007.12.15

肌肉脂肪 (intramuscular fat IMF)

是沉积在肌肉内的脂肪，是猪肉滋润多汁的物理因子，也是产生风味化合物的前体物质，是肉质测定中的重点项目之一



微
少

中

适

丰

研究对象

“华中两头乌”



Tongcheng pigs



Landrace

- 具有遗传性能稳定、繁殖力高、肉质细嫩鲜美、杂交配合力好、适应性强等突出的优良特性。

- 碳酸酐酶3（carbonic anhydrase III, CA3）主要存在于人与动物骨骼肌I型肌纤维的胞浆中，约占胞质可溶性蛋白的20%，II型肌纤维中含量甚微。
- 它也存在于包括唾液腺、输尿管平滑肌细胞、红细胞、前列腺、肺、肾、结肠和睾丸等组织中，但表达量很低

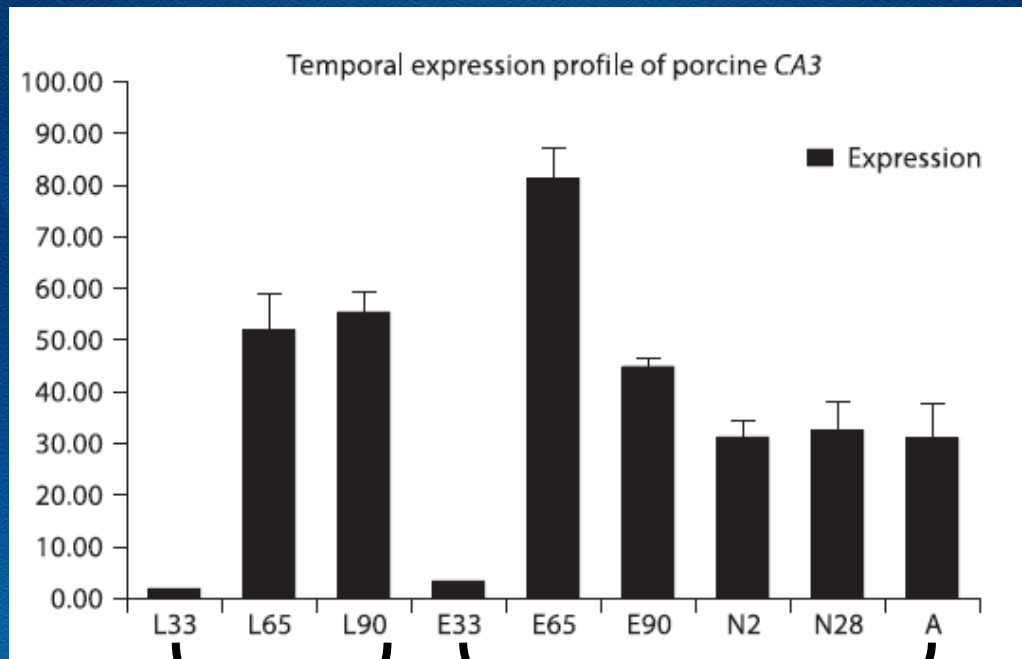
alpha-carbonic anhydrase family.

- At least 14 members have been identified in the CA family:
- CA1, CA2, CA3, CA7, and CA13 (cytoplasmic)
- CA5A and CA5B (mitochondrial)
- CA6 (secretory granules)
- CA4, CA9, CA12, and CA14 (membrane-bound),
- three related proteins: CA8, CA10, and CA11.

碳酸酐酶3的功能



- ① 由于CA3具有水化酶的功能，因此可以促进将I型肌纤维代谢过程中产生的 CO_2 运送出细胞。由于CA3能将 CO_2 运送出细胞，因此能起到调节细胞内pH的作用。
- ② 由于CA 3具有磷酸酶活性，它可能参与细胞内的信号转导。
- ③ 由于肝脏，脂肪细胞和骨骼肌的功能均与能量代谢有关，而CAIII主要存在于这些控制能量代谢的组织或细胞内，因此CA3可能具有能量代谢调节剂的作用。



Results of expression profiles showed that :

The expression levels of CA3 increased in skeletal muscles from prenatal 33- to 65-day-old Chinese Tongcheng pigs.

These levels subsequently decreased to a steady state in prenatal 90-day-old, postnatal 2-day-old, postnatal 28-day-old, and pregnant 65-day-old pigs.

SNP分析

Twelve polymorphic sites were found in the CA3 genomic DNA sequence
We analyzed the polymorphic site in position 8607 (in the sixth intron)

Table 3. Genotype and allele frequency of CA3 in different breeds

Breeds	Number	Genotype			Allele frequency	
		CC	CT	TT	C	T
Yorkshire	30	26	4	0	0.93	0.07
Landrace	29	28	1	0	0.98	0.02
Tongcheng pigs	40	2	12	26	0.20	0.80
Laiwu pigs	35	0	3	32	0.04	0.96
Guizhou mini-pigs	24	3	3	17	0.20	0.80

Statistical analysis showed the polymorphism of the CA3 gene was prominently associated with intramuscular fat content with a significance level of $P = 0.002$.



目的:

- 用所学的生物信息学知识对该蛋白的序列、结构、功能进行分析

材料

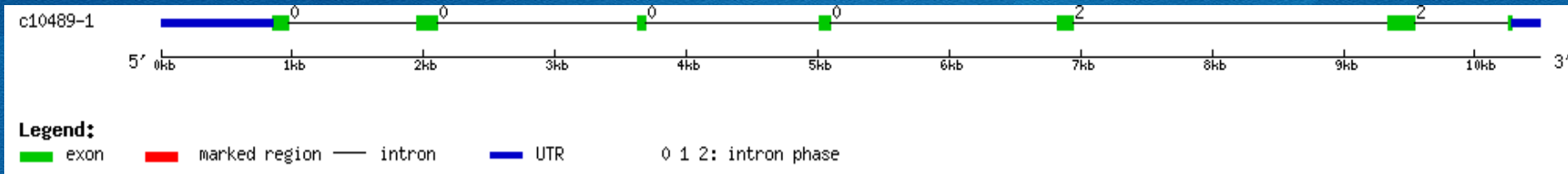
- 野猪碳酸酐酶3蛋白和核苷酸序列（来自Expasy 和NCBI）
- 碳酸酐酶家族蛋白序列（来自Expasy）

- >CAH3_PIG|Q5S1S4.3| Carbonic anhydrase 3
(Carbonic anhydrase III) (Carbonate dehydratase III)
(CA-III)

MAKEWGYADHNGPDHWHELYPIAKGDNQSPIELHTK
DIKHDPSELLPWTASYDPGSAKTILNNGKTCRVVFDDTY
DRSMLRGGPLTAAYRLRQFHLHWGSSDDHGSEHTVD
GVKYAAELHLVHWNSKYNSFATALKHPDGVAVVGIFL
KIGREKGEFQLVLDALDKIKTKGKEAPFTNFPNSCLFPA
CRDYWTYHGSFTTPPCEECIVWLLLKEPITVSSDQMA
KLRSLYSSAENEPPVPLVRNWRPPQPIKGRIVKASFK

260AA

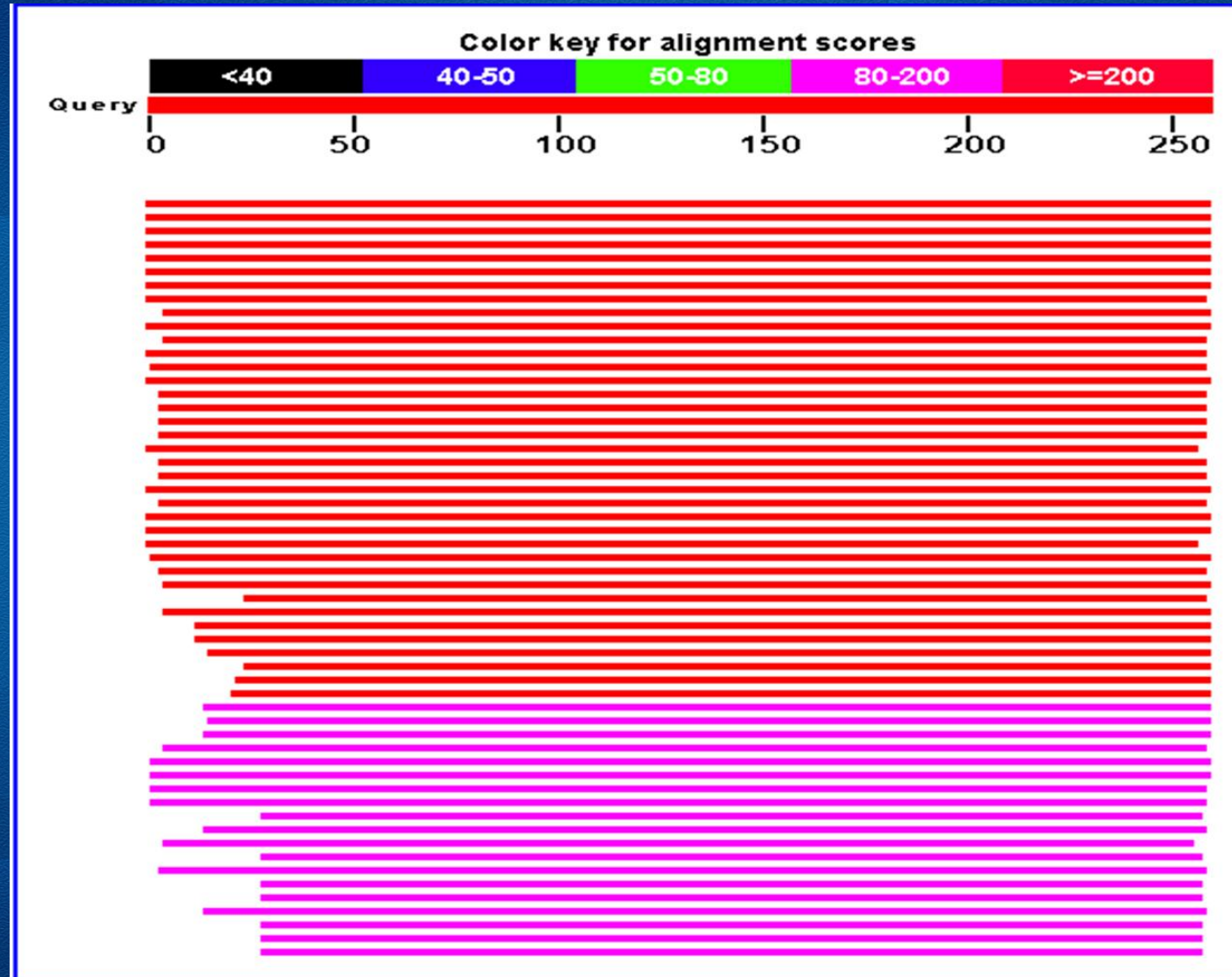
CA3 genomic DNA consists of seven exons and six introns



10.5 kb

4q11 - q14.

1、blastp



- 通过blastp可得到96条与CHA3_PIG相似的序列，大部分相似性都很高
- 91个蛋白E值从 $3e-153$ — $5e-08$
- 有2个蛋白E值为0.007和0.009
- 有3个蛋白E值大于1 (1.1,1.9,6.5)
- 所得到的相似性序列基本都是碳酸酐酶蛋白家族成员，包括：灵长类、偶蹄类、啮齿类、鸟类、有袋类、鱼、病毒、蓝藻、线虫、细菌等。

序号	Sequences producing significant alignments:	Score (Bits)	E Value
1	sp Q5S1S4 CAH3_PIG	540	3E-153
2	sp Q3SZX4 CAH3_BOVIN	510	2E-144
3	sp P16015 CAH3_MOUSE	505	8E-143
4	sp P07450 CAH3_HORSE	504	1E-142
5	sp P14141 CAH3_RAT	499	5E-141
6	sp P07451 CAH3_HUMAN	478	1E-134
7	sp P07630 CAH2_CHICK	349	5E-96
8	sp Q9D6N1 CAH13_MOUSE	330	2E-90
9	sp P00918 CAH2_HUMAN	322	8E-88

46	sp P20508 CAHH_VACCC	147	6E-35
47	sp Q10462 CAH5_CAEEL	146	8E-35
48	sp P23280 CAH6_HUMAN	145	1E-34
49	sp Q9JFA1 CAHH_VACCT	145	1E-34
50	sp P33065 CAHH_VARV	144	3E-34
51	sp QSV4Y0 CAHH_MONPV	144	3E-34
52	sp P08060 CAH6_SHEEP	144	4E-34
53	sp Q6RZ19 CAHH_RABPU	143	5E-34
54	sp Q865C0 CAH6_CANFA	143	6E-34

10	sp Q8N1Q1 CAH13_HUMAN	128	3E-29
11	sp P00919 CAH2_RABIT	127	6E-29
12	sp Q8HY33 CAH1_MONDO	124	4E-28
13	sp P00920 CAH2_MOUSE	123	9E-28
14	sp P00916 CAH1_MACMU	124	4E-28
15	sp P13634 CAH1_MOUSE	123	9E-28
16	sp P35217 CAH1_MACNE	122	2E-27
17	sp Q7M316 CAH1_GORGO	117	6E-26
18	sp P27139 CAH2_RAT	114	3E-25
19	sp P48282 CAH1_SHEEP	114	3E-25
20	sp P00921 CAH2_BOVIN	114	3E-25
21	sp Q8UWA5 CAH2_TRIHK	114	3E-25
22	sp P00915 CAH1_HUMAN	114	3E-25
23	sp Q7M317 CAH1_PANTR	114	3E-25
24	sp Q92051 CAHZ_BRARE	114	3E-25
25	sp P00917 CAH1_HORSE	114	3E-25
26	sp P00922 CAH2_SHEEP	114	3E-25
27	sp P83299 CAH1_CHIHA	114	3E-25
28	sp Q1LZA1 CAH1_BOVIN	113	6E-25
29	sp P43166 CAH7_HUMAN	113	7E-25
30	sp Q9ERQ8 CAH7_MOUSE	113	9E-25
31	sp P07452 CAH1_RABIT	113	9E-25
32	sp Q66HG6 CAH5B_RAT	113	1E-24
33	sp Q9QZA0 CAH5B_MOUSE	112	1E-24
34	sp Q9Y2D0 CAH5B_HUMAN	112	2E-24
35	sp P23589 CAH5A_MOUSE	109	1E-23
36	sp P43165 CAH5A_RAT	108	2E-23
37	sp P35218 CAH5A_HUMAN	106	1E-22
38	sp P28651 CAH8_MOUSE	97.1	7E-20
39	sp P35219 CAH8_HUMAN	96.7	9E-20
40	sp Q5PPN4 CAH8_RAT	95.9	2E-19
41	sp Q8CI85 CAH12_MOUSE	95.5	2E-19
42	sp Q9WVT6 CAH14_MOUSE	95.1	3E-19
43	sp O43570 CAH12_HUMAN	85.9	2E-16
44	sp Q9MZ30 CAH12_RABIT	80.1	8E-15
45	sp Q9ULX7 CAH14_HUMAN		

88	sp Q84UV8 NEC3_NICLS	80.1	8E-15
89	sp P54212 CAH_DUNSA	70.5	7E-12
90	sp P24258 CAH2_CHLRE	58.2	3E-08
91	sp P20507 CAH1_CHLRE	57.8	5E-08
92	sp Q8WTZ4.2 CA5BL_HUMAN	40.4	0.007
93	sp Q5SRX1 TM1L2_MOUSE	40	0.009
94	sp P47120 DOHH_YEAST	33.1	1.1
95	sp O23049 RK6_ARATH	32.3	1.9
96	sp Q6ZVM7 TM1L2_HUMAN	30.4	6.5

	_VACCV	143	8E-34
	H_VACCA	141	2E-33
	_BOVIN	140	5E-33
	3_CAEEL	140	6E-33
	5_MOUSE	137	5E-32
	10_HUMAN	135	2E-31
	9_BOVIN	132	2E-30
	6_MOUSE	130	4E-30
	_CAEEL	129	1E-29
	H_ANASP	128	3E-29
	_MOUSE	127	6E-29
	4_RABIT	124	4E-28
	_HUMAN	123	9E-28
	H4_RAT	122	2E-27
	_HUMAN	117	6E-26
	11_PONPY	116	7E-26
	1_MOUSE	114	3E-25
	114	3E-25	
	113	6E-25	
	113	7E-25	
	113	9E-25	
	113	1E-24	
	112	1E-24	
	112	2E-24	
	109	1E-23	
	108	2E-23	
	106	1E-22	
	97.1	7E-20	
	96.7	9E-20	
	95.9	2E-19	
	95.5	2E-19	
	95.1	3E-19	
	85.9	2E-16	
	80.1	8E-15	

2、SMART

- 我们发现blast到6个CHA3蛋白

序号	序列名称	结构域	序列长度	E值和结构域起始位置
1	sp Q5S1S4 CAH3_PIG	?	260	
2	sp Q3SZX4 CAH3_BOVIN		260	4.3E-177 (5-259)
3	sp P16015 CAH3_MOUSE		260	6.9E-181 (5-259)
4	sp P07450 CAH3_HORSE	?	260	
5	sp P14141 CAH3_RAT		260	5.5E-180 (5-259)
6	sp P07451 CAH3_HUMAN		260	1.4E-178 (5-259)

Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
260	1358.0	242/260 (93.1%)	253/260 (97.3%)	0/260 (0.0%)

```

Q5S1S4      1 MAKEWGYADHNGPDHWHELYPIAKGDNQSPIELHTKDIKHDPSLLPWTAS      50
      |||
Q3SZX4      1 MAKEWGYADHNGPDHWHELFPNAKGENQSPIELNTEIISHDPSLKPWTAS      50
  
```

```

Q5S1S4      51 YDPGSAKTILNNGKTCRVVFD DTYDRS
      |||
Q3SZX4      51 YDPGSAKTILNNGKTCRVVFD DTYDRS
  
```

```

Q5S1S4      101 DDHGSEHTVDGVKYAAELHLVHWN SKY
      |||
Q3SZX4      101 DDHGSEHSVDGVKYAAELHLVHWN SKY
  
```

```

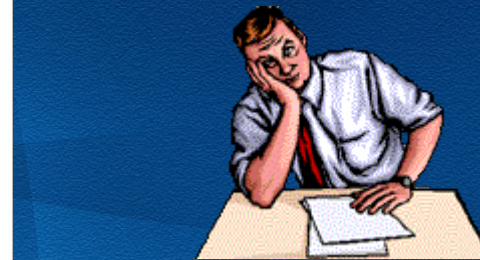
Q5S1S4      151 REKGEFQLVLDALDKIKTKGKEA PFTN
      |||
Q3SZX4      151 REKGEFQLLLDALDKIKTKGKEA PFTN
  
```

```

Q5S1S4      201 PCEECIVWLLLKEPITVSSDQMAK LRS
      |||
Q3SZX4      201 PCEECIVWLLLKEPITVSSDQIAK LRT
  
```

```

Q5S1S4      251 KGRIVKASF 260
      |||
Q3SZX4      251 KGRIVKASF 260
  
```



Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
255	1413.0	255/255 (100.0%)	255/255 (100.0%)	0/255 (0.0%)

```

Q5S1S4      5 WGYADHNGPDHWHELYPIAKGDNQSPIELHTKDIKHDPSLLPWTASYDPG      54
      |||
      1 WGYADHNGPDHWHELYPIAKGDNQSPIELHTKDIKHDPSLLPWTASYDPG      50
  
```

```

Q5S1S4      55 SAKTILNNGKTCRVVFD DTYDRSMLRGGPLTAAYRLRQFHLHWGSSDDHG      104
      |||
      51 SAKTILNNGKTCRVVFD DTYDRSMLRGGPLTAAYRLRQFHLHWGSSDDHG      100
  
```

```

Q5S1S4      105 SEHTVDGVKYAAELHLVHWN SKYNSFATALKHPDGVAVVGIFLKI GREKG      154
      |||
      101 SEHTVDGVKYAAELHLVHWN SKYNSFATALKHPDGVAVVGIFLKI GREKG      150
  
```

```

Q5S1S4      155 EFQLVLDALDKIKTKGKEA PFTNFNPSCLFPACRDYWTYHGSFTTPPCEE      204
      |||
      151 EFQLVLDALDKIKTKGKEA PFTNFNPSCLFPACRDYWTYHGSFTTPPCEE      200
  
```

```

Q5S1S4      205 CIVWLLLKEPITVSSDQMAK LRSLYSSAENEPPVPLVRNWRPPQPIKGRI      254
      |||
      201 CIVWLLLKEPITVSSDQMAK LRSLYSSAENEPPVPLVRNWRPPQPIKGRI      250
  
```

```

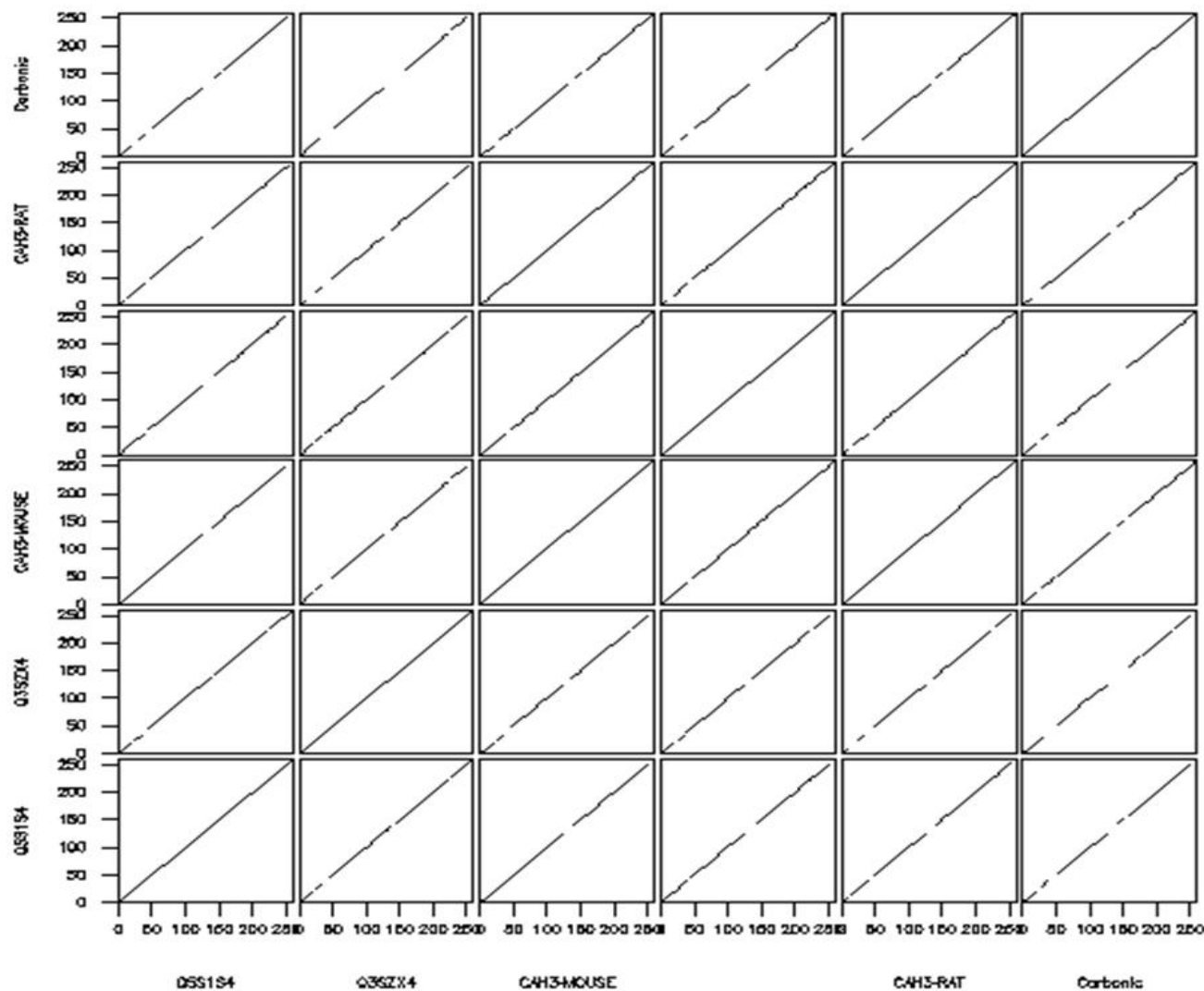
Q5S1S4      255 VKASF 259
      |||
      251 VKASF 255
  
```

CAH3_PIG CAH_BOVINE

Result: 6CHA3. View: Raw View

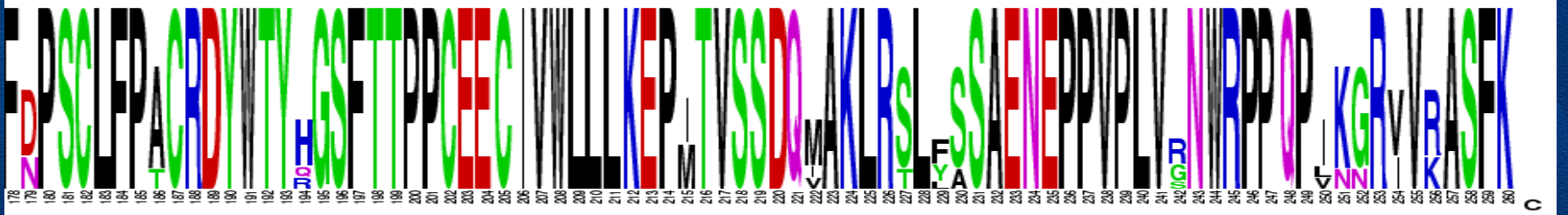
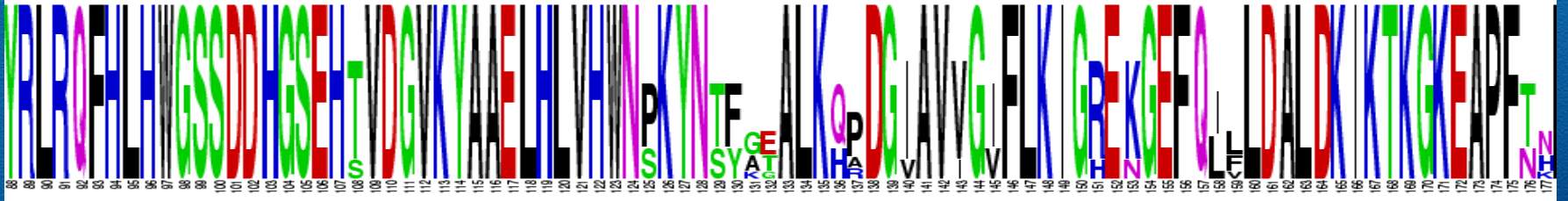
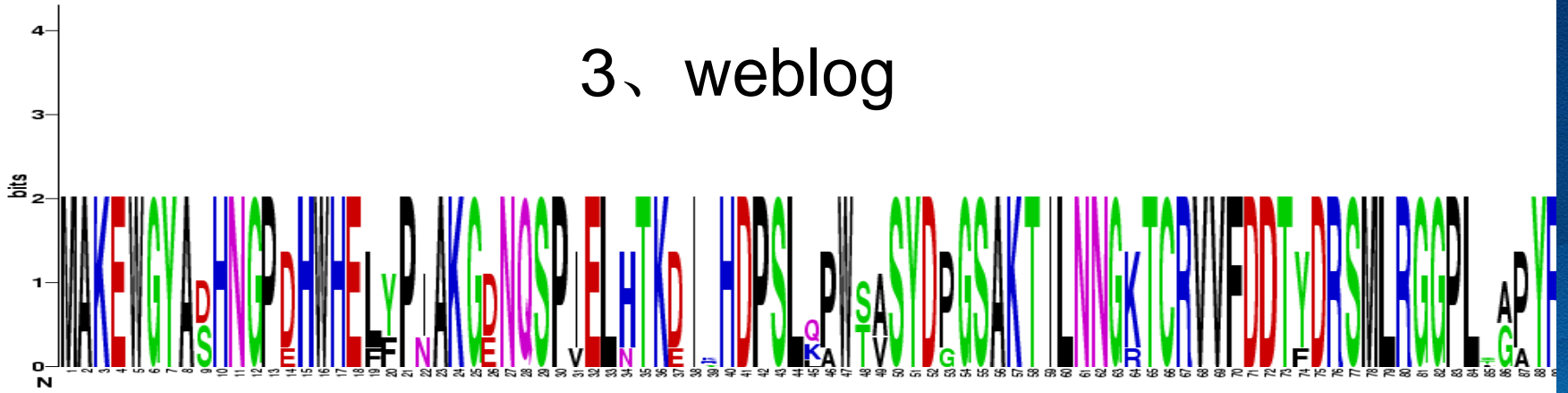
Poly dotplot of 112835

Fri 30 Nov 2007 21:08:49



No.	Length	Lines	Points	Sequence
1	260	66	1297	Q351S4
2	260	61	1230	Q35ZX4
3	260	60	1368	CAH3-MOUSE
4	260	74	1314	
5	260	62	1333	CAH3-RAT
6	260	63	1227	Carbonic

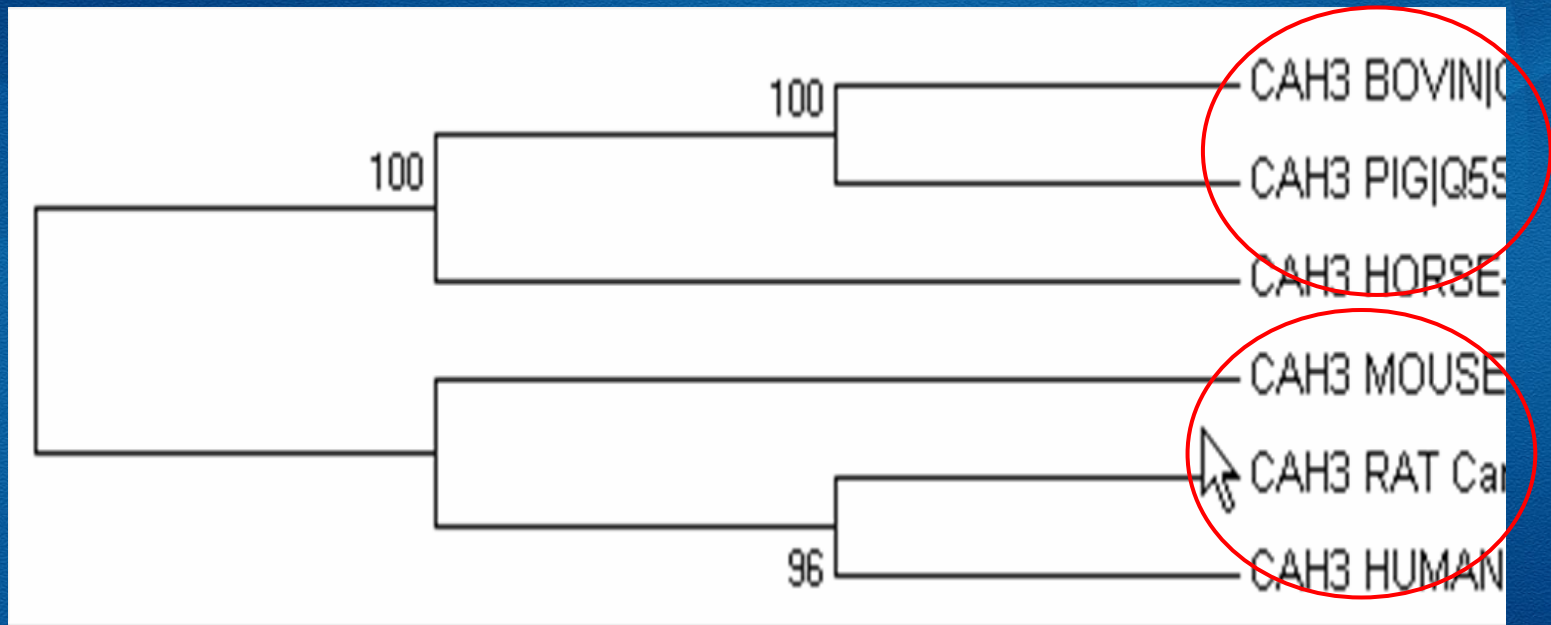
3、weblog

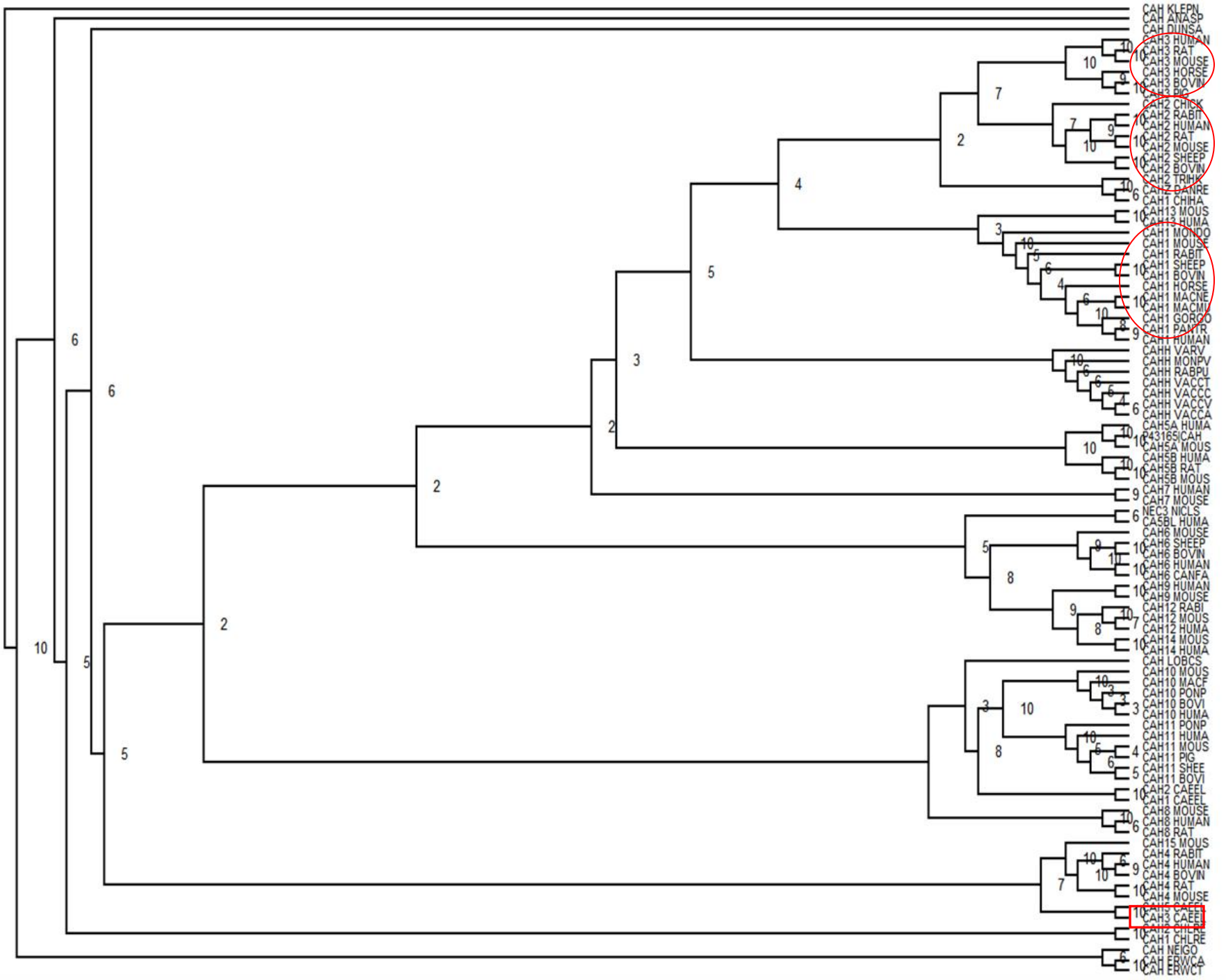


- 后来把猪和马两条一起提交还是没有 domain ?
- 最后又单独把猪和马的序列提交，找到255个残基的PFAM结构域（5—259）

序号	序列名称	结构域	序列长度	E值和结构域起始位置
1	sp Q5S1S4 CAH3_PIG		260	1.4E-180 (5-259)
2	sp Q3SZX4 CAH3_BOVIN		260	4.3E-177 (5-259)
3	sp P16015 CAH3_MOUSE		260	6.9E-181 (5-259)
4	sp P07450 CAH3_HORSE		260	3.4E-187 (5-259)
5	sp P14141 CAH3_RAT		260	5.5E-180 (5-259)
6	sp P07451 CAH3_HUMAN		260	1.4E-178 (5-259)

4、MEGA 的最大似然法构建系统发育树





5、二级结构预测 garnier分析

```

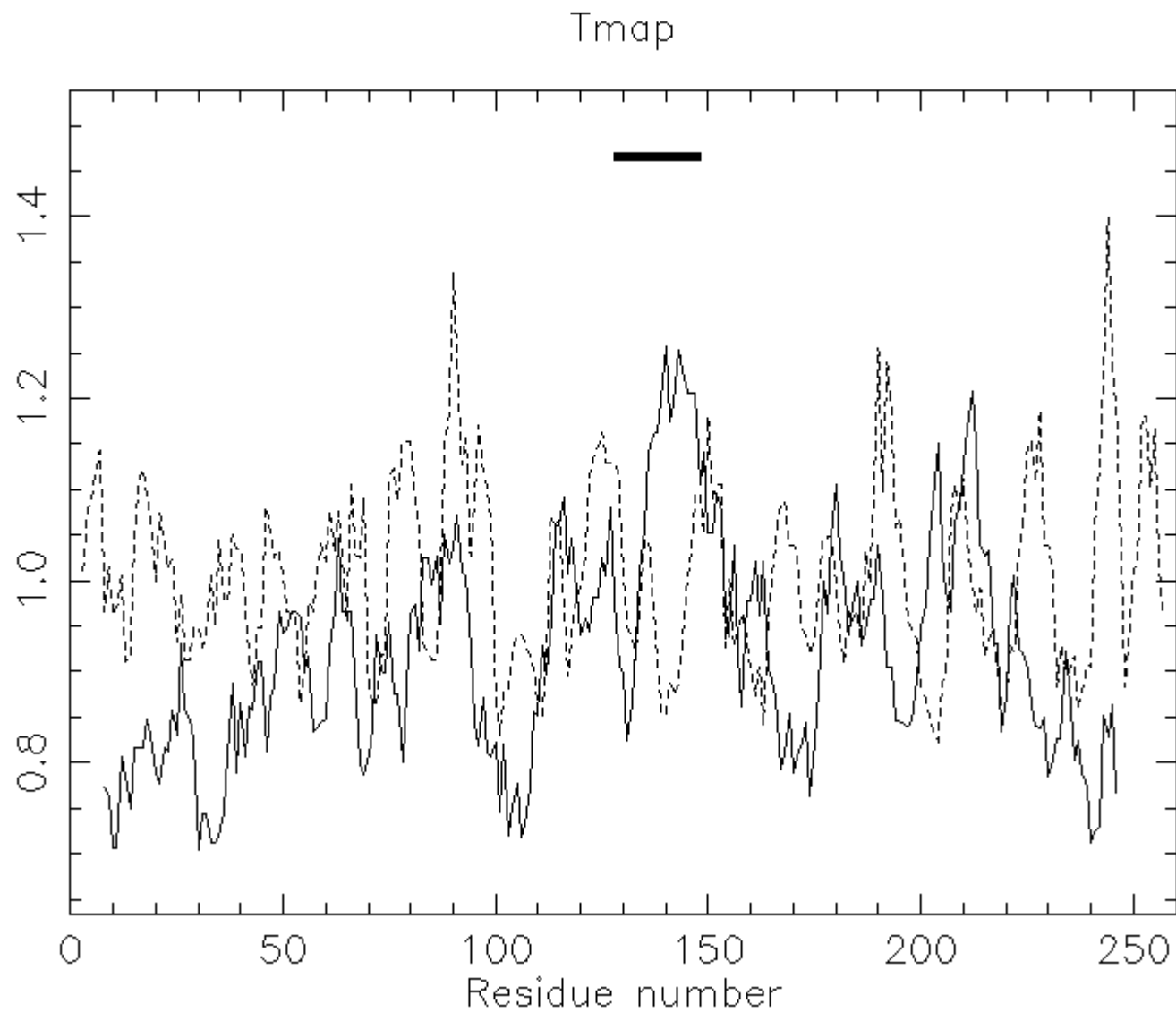
      10      20      30      40      50
MAKEWGYADHNGPDHWHELYPIAKGDNQSPIELHTKDIKHDPSSLWPWTAS
helix  HHHHHH          HH          HHHHHH
sheet
turns          T TT   TT   TTTTTT          TTT   T   TT   T
coil          C  CCC   C   CCCC          CCC CC
      60      70      80      90      100
YDPGSAKTILNNGKTCRVVFD DTYDRSMLRGGPLTAAYRLRQFHLHWGSS
helix
sheet          EEEEE   EEEEE   EE   EEE   EEE
turns          T          TTTTT   TTTTT   TT          TT
coil  CCC C          CCCC          C  CCC
      110     120     130     140     150
DDHGSEHTVDGVKYAAELHLVHWNSKYNSFATALKHPDGVAVVGVIFLKIG
helix  HHHHHHHHHHHHHHHHHHH          HHHHH          H
sheet
turns          T TTT          T
coil  CCCCC          C  CC          CCC          CC
      160     170     180     190     200
REKGEFQLVLDALDKIKTKGKEAPFTNFNPSCLFPACRDYWTYHGSFTTP
helix  H HHHHHHHHHHHHHHH
sheet          EE          EE
turns          TTTT TTTT   TTTTTTTTTTTTT
coil  CC C          CCCC          C          CC
      210     220     230     240     250
PCEECIVWLLLKEPITVSSDQMAKLRSLYSSAENEPPVPLVRNWRPPQPI
helix  HHHHHHHH          HHHHHHHHH
sheet          E          EEE
turns  TTT          TT          T          TTT   T   T

```

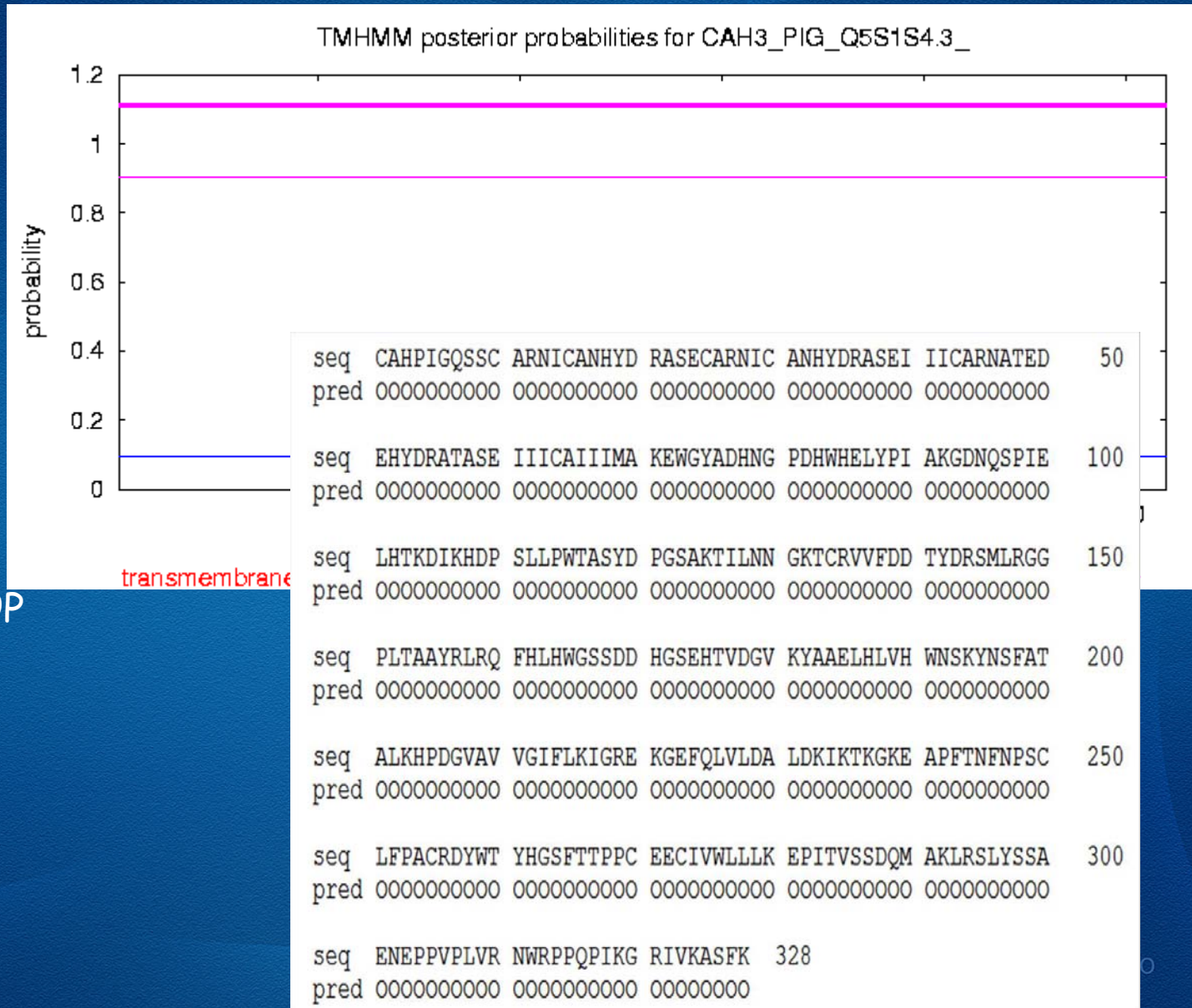
序列存在helix sheet turns coil 比例为：H: 31.1, E: 18.9, T: 29.5, C: 27.0

6、跨膜预测

Tmap



TMHMM



TMMTOP

7、3D structure

- 目前还没有猪CA3蛋白的X衍射结构

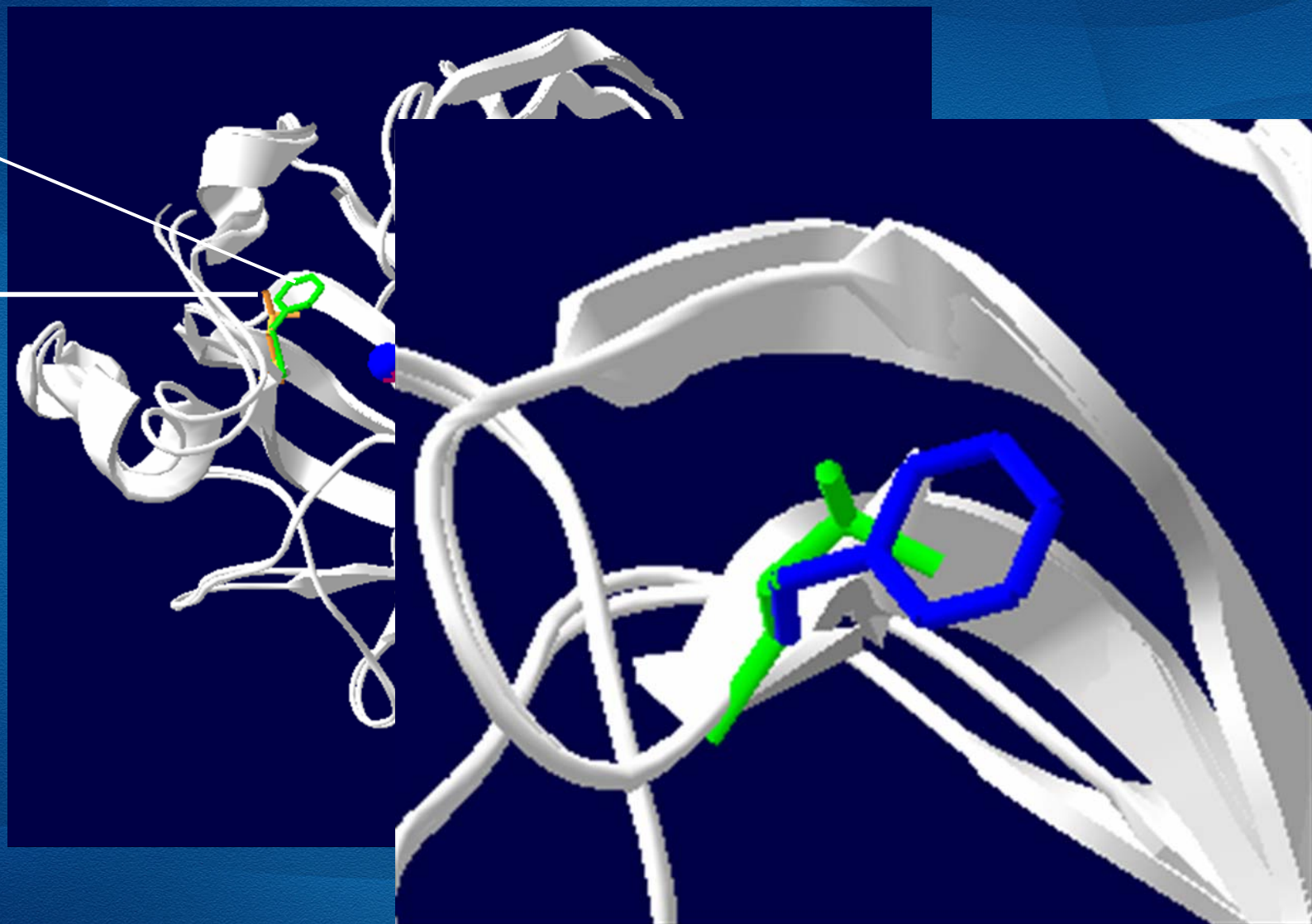


PDBsum

1Z97/1Z93 Phe-Leu(198)

PHE

LEU

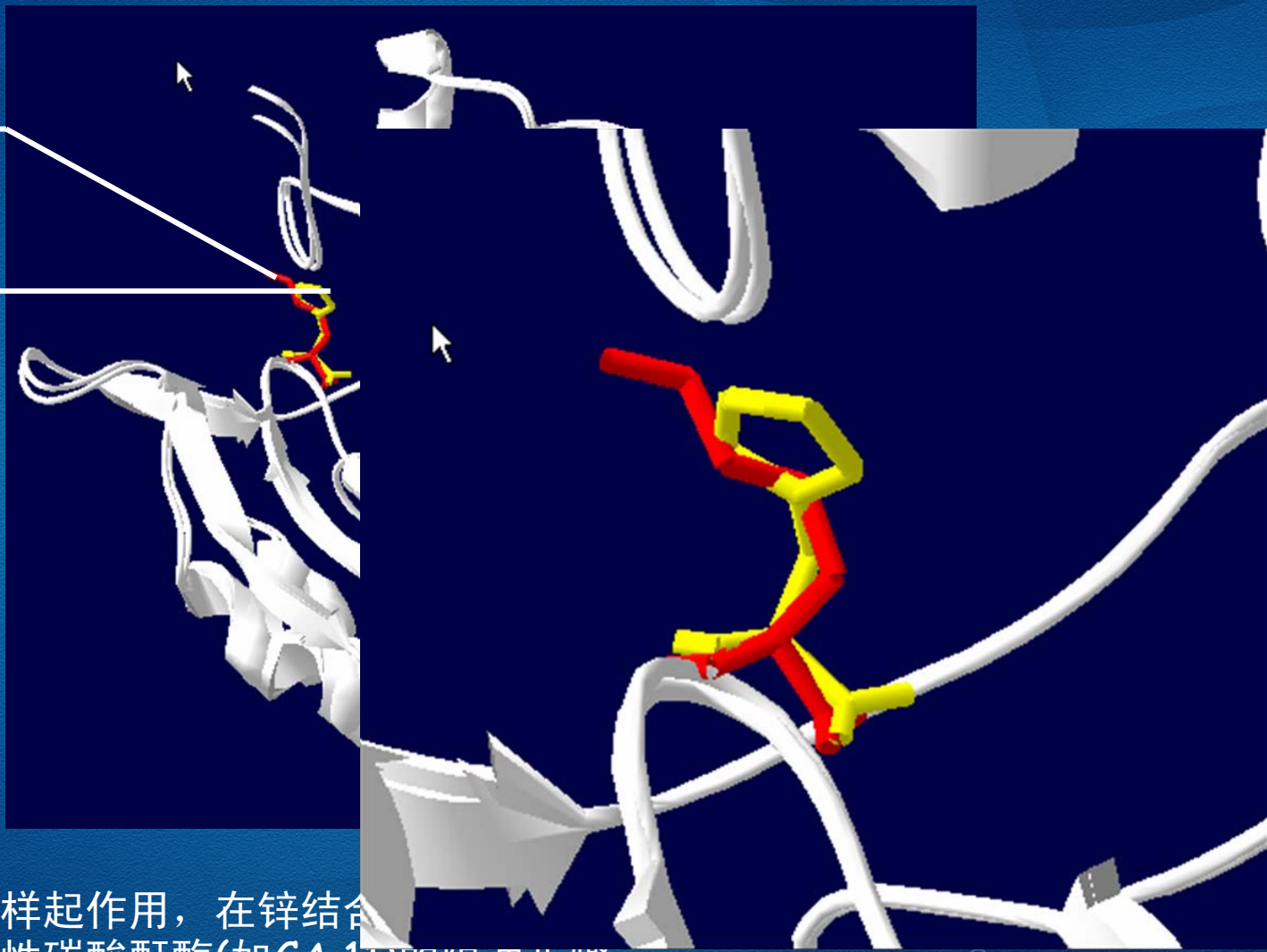


2HFX/2HFY

64:Lys-His

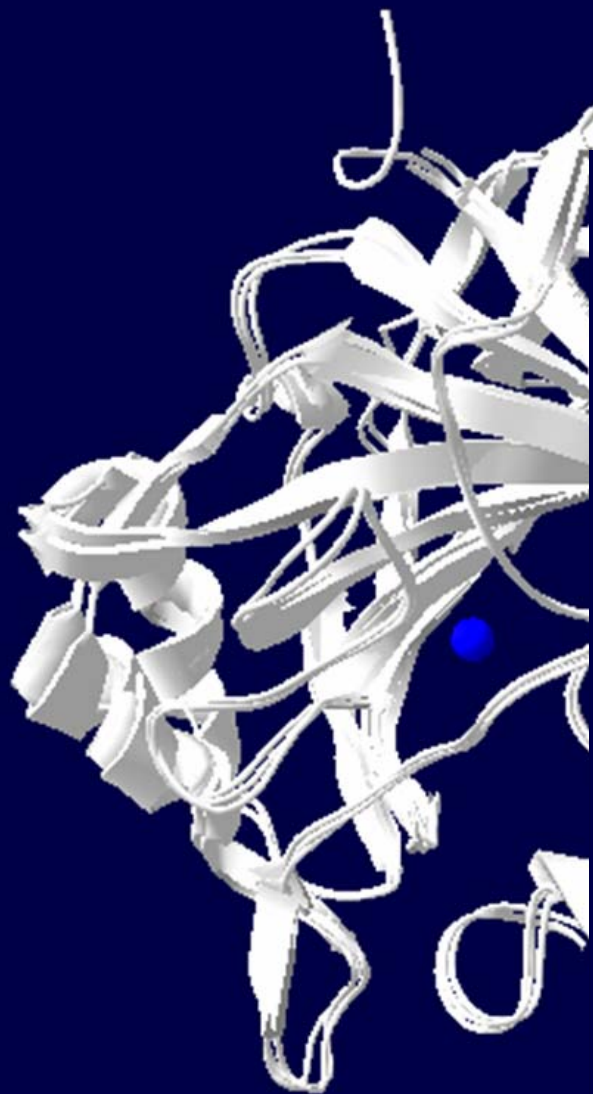
His

lys



像质子穿梭残基样起作用，在锌结合
此反应也是高活性碳酸酐酶(如CA 11)的限速步骤。

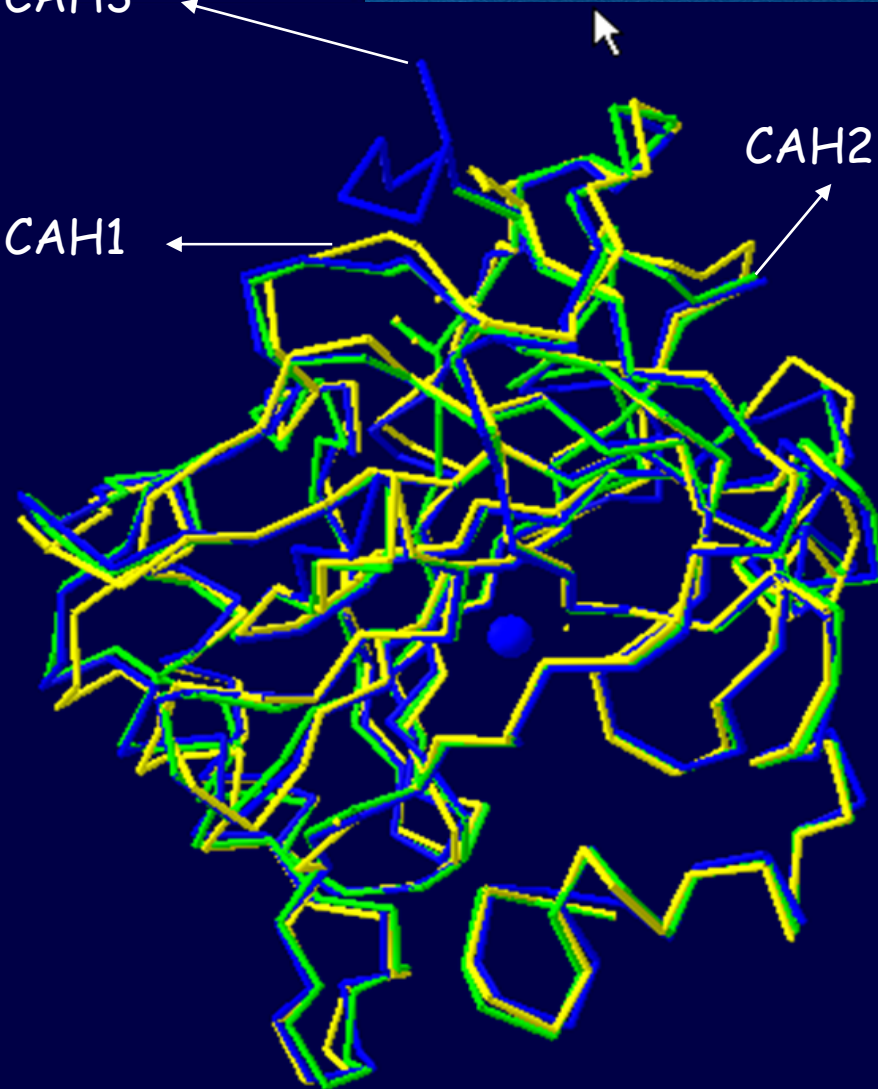
HUMAN :CAH 1,2,3
(1,2,3,7,13)

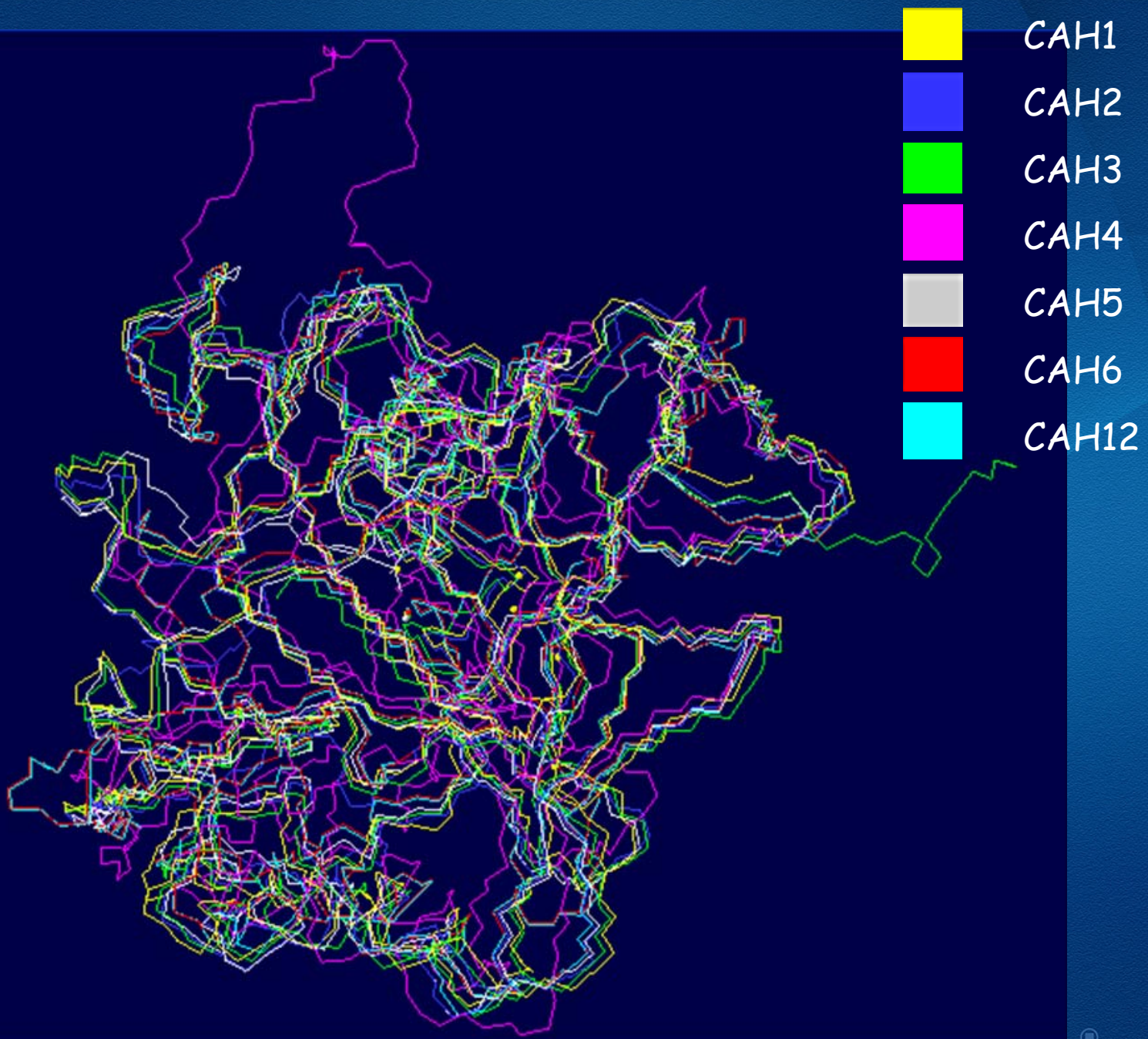


CAH3

CAH1

CAH2





- 生物信息学是帮助我们进行科学试验和研究的有力工具，但在分析时要细心、多思考，多问为什么，不能一味倚赖分析结果

Thank you !