

## 对人癌胚抗原相关蛋白质的生物信息学工具分析

水稻组：龚俊义 潘哲超 吕桂云 张春秋

人癌胚抗原是一种富含多糖的蛋白复合物，主要存在于直、结肠癌组织和胎儿肠粘膜内，属性为膜表面蛋白质。在个体发育过程中，人癌胚抗原蛋白只在胎儿时期大量表达，随着发育过程的不断深入，许多人癌胚抗原的表达明显减少甚至停滞表达。成年细胞发生癌变后会出现去分化现象，导致原先被关闭表达的人癌胚抗原再次活跃起来，重新达到胎儿时期的表达状态。基于其生理功能的重要性，研究人癌胚抗原应该具有很强烈的现实意义。

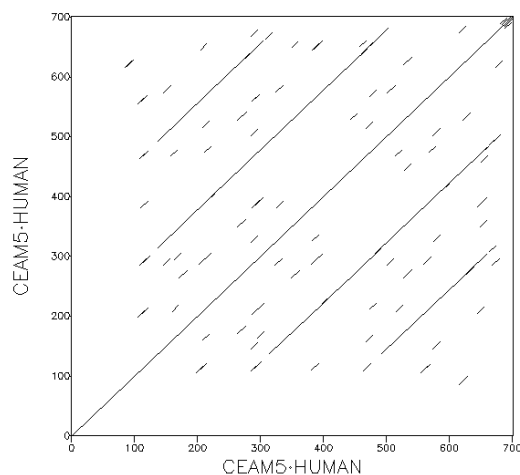
### 分析之一：人癌胚抗原相关蛋白质的序列搜索与比对

通过 ExPASy 搜索获得 17 条与关键词 “carcinoembryonic antigen” 相关的蛋白质序列，分别是：CEA16\_HUMAN (Q2WEN9)、CEA19\_HUMAN (Q7Z692)、CEA20\_HUMAN (Q6UY09)、CEA21\_HUMAN (Q3KPI0)、CEAB\_RAT (Q10753)、CEAM1\_HUMAN (P13688)、CEAM1\_MOUSE (P31809)、CEAM1\_RAT (P16573)、CEAM2\_MOUSE (Q925P2)、CEAM3\_HUMAN (P40198)、CEAM3\_RAT (Q63111)、CEAM5\_HUMAN (P06731)、CEAM6\_HUMAN (P40199)、CEAM7\_HUMAN (Q14002)、CEAM8\_HUMAN (P31997)、CEAMA\_MOUSE (Q61400) 和 PSG3\_HUMAN (Q16557)。其中 CEAB\_RAT (Q10753) 只有部分序列，CEAM5\_HUMAN (P06731) 为当前研究较为深入的蛋白质。

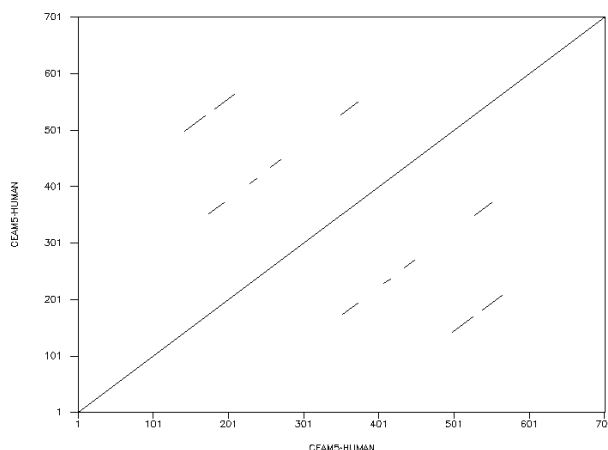
蛋白质序列比对的信息如下：

#### (1) CEAM5\_HUMAN 自身序列的 Dotmatcher 和 Dottup 比对

Dotmatcher: fasta::113167:CEAM5-HUMAN vs fasta:  
(windowsize = 10, threshold = 23.00 05/12/07)



Dottup: fasta::113187:CEAM5-HUMAN vs fasta::113188:CEAM5.  
Wed 5 Dec 2007 09:31:27

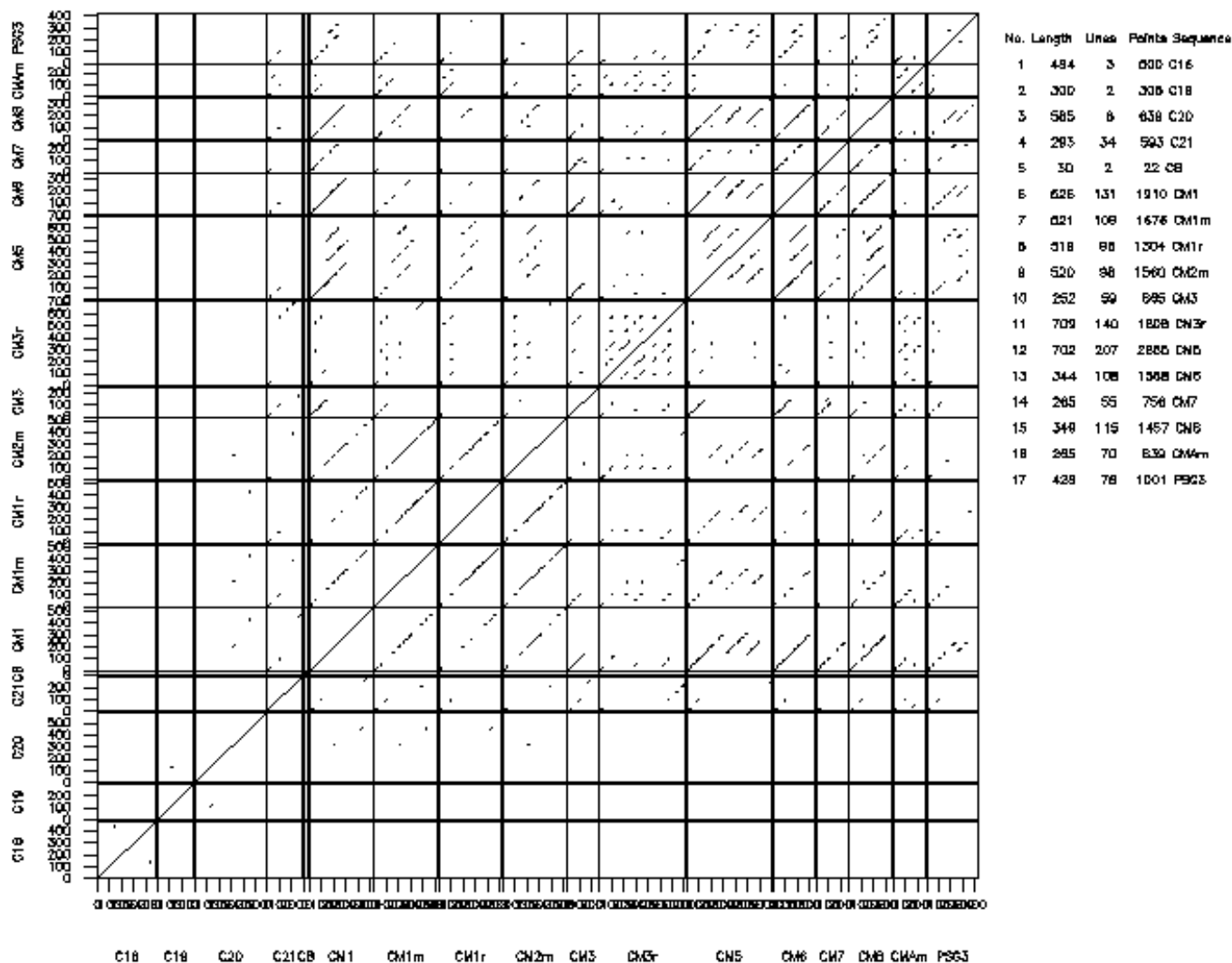


可以看出：CEAM5\_HUMAN 自身序列内部存在不同程度的序列重复，这些重复序列与后面将要研究的免疫球蛋白结构域又有如何的关联呢？需要接下来的进一步工具分析。

(2) 17 条蛋白质信息的 Poly dotplot 比对:

# Poly dotplot of 113178

Wed 5 Dec 2007 09:25:56



可以看出: CEA16\_HUMAN (Q2WEN9)、CEA19\_HUMAN (Q7Z692)和 CEA20\_HUMAN (Q6UY09)、三个序列之间以及与其它 14 个序列之间的序列差异很大, Poly dotplot 比对结果中几乎找不到重叠片段。发现序列 CEAM3\_RAT 自身比对结果中有众多杂点区, 显示出该序列存在许多小片段的重复。还发现序列 CEAM5\_HUMAN 与 CEAM1\_HUMAN、CEAM6\_HUMAN、CEAM8\_HUMAN 三个序列间存在较大片段的重复。为了看一看它们相互之间的详细序列位点差异, 我们有必要进行以下初步蛋白质信息分析和 Clustalw 全序列信息比对分析。



```

          370          380          390          400          410          420
Consensus  |-----|-----|-----|-----|-----|
CEA16_HUMAN LSGSASVVVKLSAAAVATMIVPVPTKPTREGQDVTLLTVQGYPKDLLVYAWYRGPASEPNRL
CEA19_HUMAN |-----|-----|-----|-----|-----|
CEA20_HUMAN STIEAELNSSLTLQCWAESKPGAERYRWTLLEHSTGEHLGEQLIIRALTWEHDGIYNCTASN
CEA21_HUMAN |-----|-----|-----|-----|-----|
CEAB_RAT     |-----|-----|-----|-----|-----|
CEAM1_HUMAN FKNQSLPSSERMKLSQGNNTLSINPVKREDAGTYWCEVFNPI SKNQSDPIMLNVNYNALP
CEAM1_MOUSE SQSLQLTERMTLSQNNILRIDPIKREDAGEYOCEISNPNVSVRRSNSIKLDIIFDPTQGG
CEAM1_RAT   SLQLTDRMTLSQDNSTLRIDPIKREDAGDYQCEISNPNVSVFRISHPIKLDVIPDPTQGNNG
CEAM2_MOUSE NDTLLITEKMTTSQAGLILKIDPIKREDAGEYOCEISNPNVSVKRSNSIKLEVIDFSTYDI
CEAM3_HUMAN |-----|-----|-----|-----|-----|
CEAM3_RAT   TLRLTSLDCLKAKVHVHVLQVNTSSCCDPLTPALLTIDPVPRAAAKGESVLLQVRNLPEDL
CEAM5_HUMAN QSLPVSRLQLSNDNRTLTLSSVTRNDVGPYECGIQNELSVDHSDPVIILNVLYGPDDEPTI
CEAM6_HUMAN |-----|-----|-----|-----|-----|
CEAM7_HUMAN |-----|-----|-----|-----|-----|
CEAM8_HUMAN |-----|-----|-----|-----|-----|
CEAMA_MOUSE |-----|-----|-----|-----|-----|
PSG3_HUMAN  PAEYSWTINGKFQLSGQKLFIPQITTKHSGLYACSVRNSATGMESSKSMTVKVSAVPSGTG
Consensus  |-----|-----|-----|-----|-----|

```

```

          430          440          450          460          470          480
Consensus  |-----|-----|-----|-----|-----|
CEA16_HUMAN LSQLPVSGTWIAGPAHTGREVGFPNCSSLVQKLNLTDTGRYTLKVTVTQGGKTETLEVELQV
CEA19_HUMAN |-----|-----|-----|-----|-----|
CEA20_HUMAN SLTGLARSTSVLVKVVVGPOSSSLSSGAIAGIVIGILAVIAVASELGYFLCIRNARRPSRK
CEA21_HUMAN |-----|-----|-----|-----|-----|
CEAB_RAT     |-----|-----|-----|-----|-----|
CEAM1_HUMAN QENGLSPGAIAGIVIGVVALVALIAVALACFLHFGKTGRASDQDRLTEHKPSPVSNHTQDH
CEAM1_MOUSE LSDGAIAGIVIGVVAGVALIAGLAYFLYSRKSGGGSDQDRLTEHKPSTSNHNLAPSDNSP
CEAM1_RAT   LSEGAIAGIVIGSVAGVALIAALAYFLYSRKTGGGSDHRDLTEHKPSTSSHNLGSDSDSP
CEAM2_MOUSE SDVPIAVIITGAVAGVILLAGLAYRLCSRKSRWGSQDQDRLTEHKPSSASNHNLAPSDNSFN
CEAM3_HUMAN |-----|-----|-----|-----|-----|
CEAM3_RAT   RMFIWFKSVYTSQIFKIAEYSRAINYVFRGPAHSGRETIVYTNGLLQDATEKDTGLYTL
CEAM5_HUMAN SPSYTYVRPGVNLISLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQAN
CEAM6_HUMAN |-----|-----|-----|-----|-----|
CEAM7_HUMAN |-----|-----|-----|-----|-----|
CEAM8_HUMAN |-----|-----|-----|-----|-----|
CEAMA_MOUSE |-----|-----|-----|-----|-----|
PSG3_HUMAN  HLPGLNPL
Consensus  |-----|-----|-----|-----|-----|

```

```

          490          500          510          520          530          540
Consensus  |-----|-----|-----|-----|-----|
CEA16_HUMAN x-xx-xx-x-x-xxxxx-xxxxxxxxxx-xx-xxxx-xx-x-x-xxx-x
CEA19_HUMAN APLG
CEA20_HUMAN TTEDPSHETSQPIPKKEEHPTEPSSSESLSPEYCNISQLQGRIRVELMQPPDLPEETVETKL
CEA21_HUMAN |-----|-----|-----|-----|-----|
CEAB_RAT     |-----|-----|-----|-----|-----|
CEAM1_HUMAN SNDPPNKMNEVTVSTLNFEAQOPTOPTSASPSTLATEIIVSEVKKQ
CEAM1_MOUSE NKVDDVAVTVLNFNSQCFNRPSTAPSSPRATETVYSEVKKK
CEAM1_RAT   NKVDDVSVSVLNFNAOQSKRPTSASSSPETETVYSVVKKK
CEAM2_MOUSE KVDDVAVTVLNFNSQCFNRPSTAPSSPRATETVYSEVKKK
CEAM3_HUMAN |-----|-----|-----|-----|-----|
CEAM3_RAT   QIIYRNFKIETAHVQVSVHTCVHPSTTGQLVIESVPPNVVVEGGDVLVLLVHNMPENLQSF
CEAM5_HUMAN NSASGHSRRTTVKTIIVSAELPKPSSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWVWNGQS
CEAM6_HUMAN |-----|-----|-----|-----|-----|
CEAM7_HUMAN |-----|-----|-----|-----|-----|
CEAM8_HUMAN |-----|-----|-----|-----|-----|
CEAMA_MOUSE |-----|-----|-----|-----|-----|
PSG3_HUMAN  |-----|-----|-----|-----|-----|
Consensus  |-----|-----|-----|-----|-----|

```

```

          550          560          570          580          590          600
Consensus  |-----|-----|-----|-----|-----|
CEA16_HUMAN x-x-x-x-x-x-xxxx-xxxxxxxxxx-xx-xxxx-xx-x-x-xxx-x
CEA19_HUMAN |-----|-----|-----|-----|-----|
CEA20_HUMAN PSASRRGNSFSPWPKPPKPLMPPLRLVSTVPKNMESIYEVLMGQQ
CEA21_HUMAN |-----|-----|-----|-----|-----|
CEAB_RAT     |-----|-----|-----|-----|-----|
CEAM1_HUMAN |-----|-----|-----|-----|-----|
CEAM1_MOUSE |-----|-----|-----|-----|-----|
CEAM1_RAT   |-----|-----|-----|-----|-----|
CEAM2_MOUSE |-----|-----|-----|-----|-----|
CEAM3_HUMAN |-----|-----|-----|-----|-----|
CEAM3_RAT   WYKGVAVVNRHEISRNIIASNRSTLGPASGRETIVSNGSLLLHNATEEDNGLYTLWTVN
CEAM5_HUMAN LPVSPRLQLSNGNRTLTLFNVTNRDARAYVCGIQNSVSVANRSDPVTLDVLYGPDPTIISP
CEAM6_HUMAN |-----|-----|-----|-----|-----|
CEAM7_HUMAN |-----|-----|-----|-----|-----|
CEAM8_HUMAN |-----|-----|-----|-----|-----|
CEAMA_MOUSE |-----|-----|-----|-----|-----|
PSG3_HUMAN  |-----|-----|-----|-----|-----|
Consensus  |-----|-----|-----|-----|-----|

```

```

          610          620          630          640          650          660
Consensus  |-----|-----|-----|-----|-----|
CEA16_HUMAN xx-x-x-xxxx-xx-x-xx-xx-x-x-x-x-xx-xx-xx
CEA19_HUMAN |-----|-----|-----|-----|-----|
CEA20_HUMAN |-----|-----|-----|-----|-----|
CEA21_HUMAN |-----|-----|-----|-----|-----|
CEAB_RAT     |-----|-----|-----|-----|-----|
CEAM1_HUMAN |-----|-----|-----|-----|-----|
CEAM1_MOUSE |-----|-----|-----|-----|-----|
CEAM1_RAT   |-----|-----|-----|-----|-----|
CEAM2_MOUSE |-----|-----|-----|-----|-----|
CEAM3_HUMAN |-----|-----|-----|-----|-----|
CEAM3_RAT   RHSETOGIHVHIHIYKPVQPFIRVTESSVRVKSSVVLTCISADTGTGSIQWLFNQNLR
CEAM5_HUMAN PDSSYLISGANLNLSCHSASNPSQYSWRINGIPQOHTQVLFIAKITPNNNGTYACFVSNL
CEAM6_HUMAN |-----|-----|-----|-----|-----|
CEAM7_HUMAN |-----|-----|-----|-----|-----|
CEAM8_HUMAN |-----|-----|-----|-----|-----|
CEAMA_MOUSE |-----|-----|-----|-----|-----|
PSG3_HUMAN  |-----|-----|-----|-----|-----|
Consensus  |-----|-----|-----|-----|-----|

```

```

          670          680          690          700          710          720
Consensus  x-x-x-xx-:-----:-----:-----:-----:-----:-----:
CEA16_HUMAN -----:-----:-----:-----:-----:-----:
CEA19_HUMAN -----:-----:-----:-----:-----:-----:
CEA20_HUMAN -----:-----:-----:-----:-----:-----:
CEA21_HUMAN -----:-----:-----:-----:-----:-----:
CEAB_RAT    -----:-----:-----:-----:-----:-----:
CEAM1_HUMAN -----:-----:-----:-----:-----:-----:
CEAM1_MOUSE -----:-----:-----:-----:-----:-----:
CEAM1_RAT   -----:-----:-----:-----:-----:-----:
CEAM2_MOUSE -----:-----:-----:-----:-----:-----:
CEAM3_HUMAN -----:-----:-----:-----:-----:-----:
CEAM3_RAT   -----:-----:-----:-----:-----:-----:
CEAM5_HUMAN -----:-----:-----:-----:-----:-----:
CEAM6_HUMAN -----:-----:-----:-----:-----:-----:
CEAM7_HUMAN -----:-----:-----:-----:-----:-----:
CEAM8_HUMAN -----:-----:-----:-----:-----:-----:
CEAMA_MOUSE -----:-----:-----:-----:-----:-----:
PSG3_HUMAN  -----:-----:-----:-----:-----:-----:
Consensus  x-x-x-xx-:-----:-----:-----:-----:-----:

```

Name	SeqLen	AlignLen	Gaps	GapLen	Ident	Similar	Differ	% Change	Weight	Description
CEAM1_MOUSE	CEAM1_MOUSE	521	545	7	24	159	26	336	70.825691	1.000000
CEAM2_MOUSE	CEAM2_MOUSE	520	545	7	25	155	32	333	71.559631	1.000000
CEAM1_RAT	CEAM1_RAT	519	545	8	26	154	24	341	71.743118	1.000000
CEAM5_HUMAN	CEAM5_HUMAN	702	723	6	21	190	11	501	73.720612	1.000000
CEAM6_HUMAN	CEAM6_HUMAN	344	365	6	21	193	11	140	47.123287	1.000000
CEAM8_HUMAN	CEAM8_HUMAN	349	370	6	21	192	10	147	48.108109	1.000000
CEAM1_HUMAN	CEAM1_HUMAN	526	545	6	19	196	12	318	64.036697	1.000000
CEAM7_HUMAN	CEAM7_HUMAN	265	273	3	8	144	17	104	47.252747	1.000000
PSG3_HUMAN	PSG3_HUMAN	428	435	3	7	130	36	262	70.114944	1.000000
CEA21_HUMAN	CEA21_HUMAN	293	303	4	10	111	26	156	63.366337	1.000000
CEAM3_HUMAN	CEAM3_HUMAN	252	264	5	12	100	12	140	62.121212	1.000000
CEAM3_RAT	CEAM3_RAT	709	715	3	6	76	33	600	89.370628	1.000000
CEAMA_MOUSE	CEAMA_MOUSE	265	271	3	6	78	23	164	71.217712	1.000000
CEA16_HUMAN	CEA16_HUMAN	484	494	5	10	81	47	356	83.603241	1.000000
CEA20_HUMAN	CEA20_HUMAN	585	589	3	4	70	37	478	88.115448	1.000000
CEA19_HUMAN	CEA19_HUMAN	300	305	2	5	34	40	226	88.852463	1.000000
CEAB_RAT	CEAB_RAT	30	30	0	0	6	1	23	80.000000	1.000000

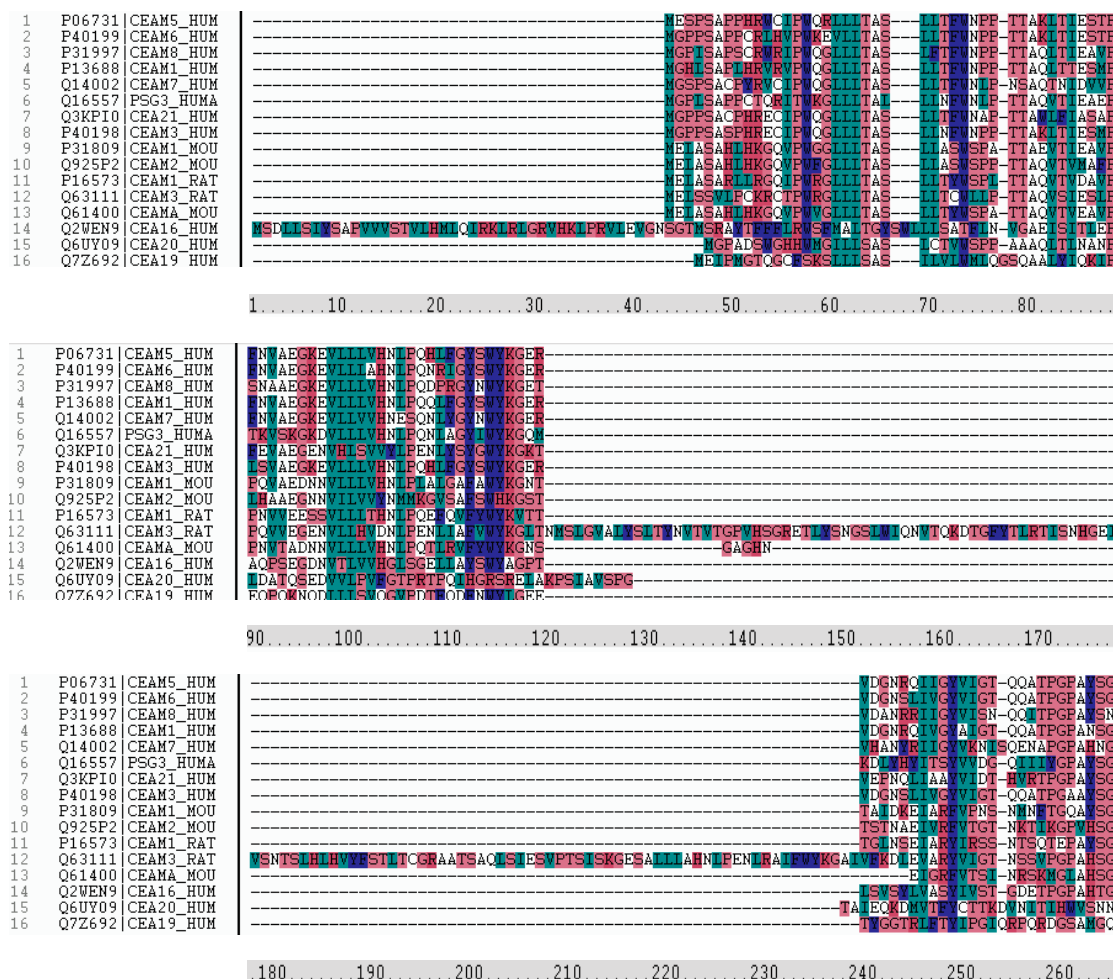
(4) Needle 和 Water 分析 17 条蛋白质信息:

Alignment pairs	Alignment types	length	scope	identity	similarity	gaps
CEAM5 & CEA16	needle	772	450.5	18.5%	28.1%	46.4%
	water	401	474.5	31.4%	46.9%	8.7%
CEAM5 & CEA19	needle	723	148.5	11.8%	19.6%	61.4%
	water	395	157	20.3%	33.9%	34.7%
CEAM5 & CEA20	needle	798	587	21.6%	33.6%	38.7%
	water	521	601	29.2%	45.5%	18.2%
CEAM5 & CEA21	needle	722	551	17.6%	22.7%	62.2%
	water	254	564.5	46.5%	59.1%	6.7%
CEAM5 & CEAM1	needle	722	1604.5	45.6%	52.2%	29.9%
	water	520	1612	62.7%	71.7%	5.0%
CEAM5 & CEAM1-MO	needle	715	1025	33.3%	43.8%	29%
	water	555	1028.5	42.2%	54.8%	11.4%
CEAM5 & CEAM1-RA	needle	706	974	32.2%	43.5%	27.1%
	water	567	975	40%	54.1%	9.5%
CEAM5 & CEAM2-MO	needle	720	995.5	31.8%	43.1%	30.3%
	water	560	999	40.2%	53.8%	13.2%
CEAM5 & CEAM3	needle	741	623	18.8%	20.9%	71.3%
	water	154	639.5	80.5%	84.4%	2.6%
CEAM5 & CEAM3-RA	needle	904	500.5	21.2%	30.5%	43.9%
	water	792	503.5	24.2%	34.8%	36.2%
CEAM5 & CEAM6	needle	702	1432	39%	40.9%	51%
	water	323	1443	83.9%	87.3%	0%
CEAM5 & CEAM7	needle	711	825	23.2%	26.9%	64%
	water	245	832.5	65.7%	75.5%	2.9%

CEAM5&CEAM8	needle	707	1255.5	34.9%	38.3%	51.3%
	water	314	1272	76.8%	82.8%	0%
CEAM5&CEAMA-MO	needle	704	343.5	14.6%	20.3%	62.6%
	water	297	352	33%	45.8%	19.2%
CEAM5&CEAPSG	needle	711	1135.5	33.6%	40.8%	41.1%
	water	496	1143	47.4%	57.7%	17.5%

可以看出，就全局比对而言，CEAM1\_HUMAN 与 CEAM5\_HUMAN 的序列相似性最高，为 52.2%；就局部比对而言，CEAM1\_HUMAN、CEAM3\_HUMAN、CEAM6\_HUMAN、CEAM7\_HUMAN、CEAM8\_HUMAN 都与 CEAM5\_HUMAN 存在较高的序列相似性。就与 CEAM5\_HUMAN 的整体比对效果而言，CEAM1\_HUMAN 综合成绩应该是最好的。同时我们还可以发现，尽管人和大、小鼠的形态差异较大，但是 CEAM1\_MOUSE、CEAM1\_RAT 和 CEAM2\_MOUSE 三个蛋白质序列与人的 CEAM5\_HUMAN 也存在较高的序列相似性。

(4) Clustalw 全序列分析 16 条蛋白质（除了其中的 CEAB\_RAT）信息：



1	P06731	CEAM5_HUM	REI...PNASLLIQN...IQNDTG...TLHV...KSD...NEEATG...QRV...P
2	P40199	CEAM6_HUM	RET...PNASLLIQN...IQNDTG...TLQV...KSD...NEEATG...QRV...P
3	P31997	CEAM8_HUM	RET...PNASLLIRN...TRNDTGS...TLQV...KSNL...NEEATG...QSV...HP
4	P13688	CEAM1_HUM	RET...PNASLLIQN...IQNDTG...TLQV...KSD...NEEATG...QRV...P
5	Q14002	CEAM7_HUM	RET...PNGTLLIQN...TRNDAG...TLHV...KENL...NEEATR...QV...VS
6	Q16557	PSG3_HUMA	RET...PNASLLIQN...TRNDAGS...TLHV...KRGD...GIRGETGH...IT...L
7	Q3KPI0	CEA21_HUM	RET...SPSGDLIQN...TIEDTG...TLQV...TRNS...EQASH...HRV...P
8	P40198	CEAM3_HUM	RET...PNASLLIQN...IQNDTG...TLQV...KSD...NEEATG...QRV...P
9	P31809	CEAM1_MOU	RET...SNGLIQN...TKKNG...TID...IDEN...RRTOAT...RH...HFL...LKPN...TSNNS...NPF...EGDSS...VSL...TCD...S...TDEP
10	Q925P2	CEAM2_MOU	RET...SNGLIQR...TKKDTG...TID...IDEN...RRR...TG...QH...HKL...KSN...TSNNS...NPF...EGDSS...VSL...TCD...S...TDEP
11	P16573	CEAM1_RAT	RET...SNGLIQN...NKTDG...TID...IDEN...RRR...TG...QH...HKL...KSN...TSNNS...NPF...EGDSS...VSL...TCD...S...TDEP
12	Q63111	CEAM3_RAT	RET...SNGLIQN...TRNDAG...TKT...SID...KTE...IA...V...Q...D...T...C...H...S...A...G
13	Q61400	CEAMA_MOU	RET...SNGLIQN...TKNDEG...TID...IDEN...E...TE...S...R...H...H...P...S...L...P...S...S
14	Q2WEN9	CEA16_HUM	REA...RPDGS...ID...QG...L...PRH...SGT...IQ...TN...Q...Q...TE...G...G...H...C...H...E
15	Q6UY09	CEA20_HUM	LSV...PHER...IQ...SKD...CK...L...L...V...Q...RED...SGT...Q...CEA...R...DA...L...S...Q...R...S...D...P...L...D...K...G...F...D...E...V...E...K...I...E...S...G...V...A...S...G
16	Q7Z692	CEA19_HUM	RDIV...G...P...N...G...S...L...L...R...R...A...Q...P...T...S...G...T...Q...M...A...L...I...N...S...E...T...M...K...A...K...T...E...Q...A...E...K...N...K...E...I...P...S

... 270 ..... 280 ..... 290 ..... 300 ..... 310 ..... 320 ..... 330 ..... 340 ..... 350 .....

1	P06731	CEAM5_HUM	-----
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	-----
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	NINVLVSRNGESTSEGDRKISEGNRTLLINVRNDTGFVCETRNPFVSNRSDPFSINIIIGPDTPIISPSDVIHPGNSNINISCH
11	P16573	CEAM1_RAT	NITVLSRNGESTSEGDRKISEGNRTLLINVRNDTGFVCETRNPFVSNRSDPFSINIIIGPDTPIISPSDVIHPGNSNINISCH
12	Q63111	CEAM3_RAT	NISVLSRNGESTSEGDRVITSEGNRTLLINVRRTDKGVECEARNEATINRSDPFLNIDVIGPDAPVISPDPVTHQGSNINISCH
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	-----
15	Q6UY09	CEA20_HUM	-----
16	Q7Z692	CEA19_HUM	-----

EVVEVMEGSSNITIAET

... 360 ..... 370 ..... 380 ..... 390 ..... 400 ..... 410 ..... 420 ..... 430 ..... 440 .....

1	P06731	CEAM5_HUM	-----
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	-----
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	AASNPPAQLFVLEKPHASSQELFIPNITINNSGTVCLVNNSVTGLSRTIVNNTVLEFVTOFELQVFNITVKEIDSDITLTCSE
11	P16573	CEAM1_RAT	AASNPPAQLFVLEKPHASSQELFIPNITINNSGTVCLVNNSVTGLSRTIVNNTVLEFVTOFELQVFNITVKEIDSDITLTCSE
12	Q63111	CEAM3_RAT	ADSNPPAQLFVLEKPHASSQELFIPNITINNSGTVCLVNNSVTGLSRTIVNNTVLEFVTOFELQVFNITVKEIDSDITLTCSE
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	-----
15	Q6UY09	CEA20_HUM	-----
16	Q7Z692	CEA19_HUM	-----

ELPKPSI SSNNSNPFEDKDAIAATCEE  
 ELPKPSI SSNNSNPFEDKDAIAATCEE  
 ELPKPSI SSNNSNPFEDKDAIAATCEE  
 ELPKPSI SSNNSNPFEDKDAIAATCEE  
 ELPKPSI SSNNSNPFEDKDAIAATCEE  
 EPPKPSI SSNNSNPFEDKDAIAATCEE  
 ELPKPSI SSNNSNPFEDKDAIAATCEE  
 ESNAQPSI QASSITVTEKGSVWLTQNI  
 QENA-FCDFVGA-NAGVIG-VLVGAA  
 AASNPPAQLFVLEKPHASSQELFIPNITINNSGTVCLVNNSVTGLSRTIVNNTVLEFVTOFELQVFNITVKEIDSDITLTCSE  
 AASNPPAQLFVLEKPHASSQELFIPNITINNSGTVCLVNNSVTGLSRTIVNNTVLEFVTOFELQVFNITVKEIDSDITLTCSE  
 ADSNPPAQLFVLEKPHASSQELFIPNITINNSGTVCLVNNSVTGLSRTIVNNTVLEFVTOFELQVFNITVKEIDSDITLTCSE  
 PPTSAQITNESAPTSVAGASVLLLVHNI  
 PPTGQITNEAPFNVAESENVLLLVHNI  
 IIAQPTVIANSTALWERDITLRLMCSSE  
 KSHPPCAMTWELDGLSHITTRITTHANSRPEHGLRCLVNSATHSSIGTLKRVVLETLINPQVVFSSNINENARSMDITCOI  
 THLPTNAGITAAITIGSAAAGALLTSC

... 450 ..... 460 ..... 470 ..... 480 ..... 490 ..... 500 ..... 510 ..... 520 .....

1	P06731	CEAM5_HUM	ETQDATTLVWNNQSPSPRIQISNGNRTLLINVRNDTASVCEITONEVSAARRSD
2	P40199	CEAM6_HUM	EQNTITLVWNNQSPSPRIQISNGNRTLLISVKNRNDAGS ECEQNPASARRSD
3	P31997	CEAM8_HUM	EQNTITLVWNNQSPSPRIQISNGNRTLLISVTRNDGCP ECEQNPASARRSD
4	P13688	CEAM1_HUM	ETQDATTLVWNNQSPSPRIQISNGNRTLLISVTRNDTGP ECEQNPASARRSD
5	Q14002	CEAM7_HUM	EQNTITLVWNNQSPSPRIQISNGNRTLLISVTRNDGCP ECEQNPASARRSD
6	Q16557	PSG3_HUMA	ETPDAGLVWNNQSPSPRIQISNGNRTLLISVTRNDGCP ECEQNPASARRSD
7	Q3KPI0	CEA21_HUM	NTNIGTSQVLENNORIQTKRMLSNRNLITDPIQEDAGEVCEQSNPSSARRSD
8	P40198	CEAM3_HUM	AAALCGLAKIGRTSQRDLKE--QDFQALAPGR--GPHSSS
9	P31809	CEAM1_MOU	NDGANQLVNSQSLQTERNTISQNNSEFIDPKREDAGEVCEQSNPSSARRSN
10	Q925P2	CEAM2_MOU	KDRQAEHLINNDTITKXHTTSQAGLTKDPIKREDAGEVCEQSNPSSARRSN
11	P16573	CEAM1_RAT	KDTGSEVLENSQSLQTERNTISQNNSEFIDPKREDAGEVCEQSNPSSARRSN
12	Q63111	CEAM3_RAT	PENIARVNRKGLKIDLEARYVGTNSSLGPAHSGREITSNGLLIQNTRNDAG
13	Q61400	CEAMA_MOU	PRTIARVNRGTTAGERNEARVITASNKLLGPAHSDREITSNGLSEVQTKNDEG
14	Q2WEN9	CEA16_HUM	PSPTAEVNRNGGALPVALRIGSPDGRVARRGIFREEAGA VCEQSNPSSARRSE
15	Q6UY09	CEA20_HUM	VNCSNQLVSGQFLPSEHLQISADNRTLLHGQRNDTGFACEVNLGSRARSE
16	Q7Z692	CEA19_HUM	IALLVTRNIRGQSHRIPAPRGSSSLCSAIVSPVPSIPSTHATTEKPEGPAHDAGD

530 ..... 540 ..... 550 ..... 560 ..... 570 ..... 580 ..... 590 ..... 600 ..... 610 .....

1	P06731	CEAM5_HUM	-----
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	-----
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	-----
11	P16573	CEAM1_RAT	-----
12	Q63111	CEAM3_RAT	-----
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	-----
15	Q6UY09	CEA20_HUM	-----
16	Q7Z692	CEA19_HUM	-----

SVLNLVGPDPITLSPINISIR  
 FVTLNVLGPDVPTLSPSMANIR  
 FVTLNVLGPDAPITLSPSDIYIR  
 FVTLNVTGPDPTLSPSDIYIR  
 FVTLNVRRES  
 FVTLNVLGPDVPTLSPINISIR  
 FLKTIK--SDDNITL  
 ASMSPL--SSAQALP  
 SKTDITLDPITGG--LSDGAA  
 GKTEVLDSTV--LSDVFLA  
 FIKDVLDPDITQNSGISEGAA  
 VVTLRTSIDAKAVVHWQVQNN  
 AADMLQNDHILPQVQNN  
 FNLIVGPERVALQDSTIRTCG  
 FLETINLGPDCVHITRESAENIS  
 NNIVEMPSFVLLVSPISDTRSN

... 620 ..... 630 ..... 640 ..... 650 ..... 660 ..... 670 ..... 680 ..... 690 ..... 700 .....

1	P06731	CEAM5_HUM	---SGENILISCHAAASNPPAQLSWLVNGTQOSTQELIPNITVNNSSGSLCOAHNSDTGKRRITVITIVAEPPKPIITSNNS
2	P40199	CEAM6_HUM	---PGENILISCHAAASNPPAQLSWLVNGTQOSTQELIPNITVNNSSGSLCOAHNSGATGKRRITVITIVAEPPKPIITSNNS
3	P31997	CEAM8_HUM	---AGNINISCHAAASNPPSQISLVNGTQQTQQLIPNITVNNSSGSLACHITNSGATGRRITVITIVAEPPKPIITSNNS
4	P13688	CEAM1_HUM	---PGANILISCHAAASNPPAQLSWLVNGTQOSTQELIPNITVNNSSGSLCOAHNSDTGKRRITVITIVAEPPKPIITSNNS
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	---SGENILISCHADSPPAELSWLVNGTQOSTQELIPNITVNNSSGSLCOAHNSDTGKRRITVITIVAEPPKPIITSNNS
7	Q3KPI0	CEA21_HUM	---GILGVLVGSLLVAALVCLLRKTRGASDOSDRREQOPFASTPGHGSDSSLS
8	P40198	CEAM3_HUM	---LNPRTAASVEELKHDITNVRNDHKAELAS
9	P31809	CEAM1_MOU	---GIVGVVAGVAVAGAVAVLVRKSGGSDORDTEHKPSTSNHNPASDNPENKDDVAIVIN
10	Q925P2	CEAM2_MOU	---VITGAVAGVAVAGAVAVLVRKSGGSDORDTEHKPSTSNHNPASDNPENKDDVAIVIN
11	P16573	CEAM1_RAT	---GIVGVVAGVAVAGAVAVLVRKSGGSDORDTEHKPSTSNHNPASDNPENKDDVAIVIN
12	Q63111	CEAM3_RAT	---TSSCCDPELPAALLIDVPRHAAKGESVLLQVNRLEEDLRNFVLRKSVITQILKIAEISRAINIVRGEAHSCHETVITING
13	Q61400	CEAMA_MOU	---HAKKQ
14	Q2WEN9	CEA16_HUM	---IILKVDNTSLILVCSRSCEPEPEVITINGQALKNGQDHNSSTAAACEGITCAAKNTKILSGSASVVKISAAAATHIVVPE
15	Q6UY09	CEA20_HUM	---TTEAEINSSITLQCAESKPGAEVITIEHSTGELIGECILTRALTIEDGLNCTASNSITCARSTVIVKVVVGCSSISSSGA
16	Q7Z692	CEA19_HUM	---PARPLTPPHQAEENHQQODLNEDPAFCOLVPTS

.....710.....720.....730.....740.....750.....760.....770.....780.....790

1	P06731	CEAM5_HUM	NEVEDEDAVAITCEPEIQNTITLVVWVNNQSLPWSRPLQISNDNRTITLISVTRNDVGFECQIQNEISVDSHSDPVLINVLVGGDDPET
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	-----
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	-----
11	P16573	CEAM1_RAT	-----
12	Q63111	CEAM3_RAT	---SLILQDATERKDTG
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	---TKPTEGQDWTILVCG
15	Q6UY09	CEA20_HUM	---AGIVIGITAVITANASEIG
16	Q7Z692	CEA19_HUM	-----

.....800.....810.....820.....830.....840.....850.....860.....870

1	P06731	CEAM5_HUM	LSPSVTVVIRPGVNSISCHAAASNPPAQLSWLVLDGNLQCHTQELIPNITVNNSSGSLCOAHNSAGHSRTIVKTIITVSAELPKPSISS
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	-----
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	-----
11	P16573	CEAM1_RAT	-----
12	Q63111	CEAM3_RAT	---LNTLQILVRRNKIEEAVHCVSWHTCQHPSTTGQLVIVDS
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	-----
15	Q6UY09	CEA20_HUM	-----
16	Q7Z692	CEA19_HUM	-----

880.....890.....900.....910.....920.....930.....940.....950.....960

1	P06731	CEAM5_HUM	SKFVEDKDAVAITCEPEAQNTITLVVWVNNQSLPWSRPLQISNGNRTITLISVTRNDVAFVCGIQNSVANSRDPVITLVDVVGDDPET
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	-----
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	-----
11	P16573	CEAM1_RAT	-----
12	Q63111	CEAM3_RAT	---PNVDPGGVLLLVHNPENIQCSLVKGVAVLVRRHESRNLIASNRSTLGFASGRETITVINGSLLLHNAITEEDNGVITLITVNRHSE
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	-----
15	Q6UY09	CEA20_HUM	-----
16	Q7Z692	CEA19_HUM	---VPLCLRNRARRPSR

970.....980.....990.....1000.....1010.....1020.....1030.....1040.....1050

1	P06731	CEAM5_HUM	LSPPDSSVLSGANILISCHAAASNPPSQISLVNRIHQHTQELIPNITVNNSSGSLCOAHNSAGHSRTIVKTIITVSAELPKPSISS
2	P40199	CEAM6_HUM	-----
3	P31997	CEAM8_HUM	-----
4	P13688	CEAM1_HUM	---ENGLSPGAAAGIVIGVAVAVAGAVAVLVRKSGGSDORDTEHKPSTSNHNPASDNPENKDDVAIVIN
5	Q14002	CEAM7_HUM	-----
6	Q16557	PSG3_HUMA	-----
7	Q3KPI0	CEA21_HUM	-----
8	P40198	CEAM3_HUM	-----
9	P31809	CEAM1_MOU	-----
10	Q925P2	CEAM2_MOU	-----
11	P16573	CEAM1_RAT	-----
12	Q63111	CEAM3_RAT	---TQGLHVHITHVYKFAQPIRIVTESSVWVYKSSVVIICISADTIGTICQLLNQNLRIITQVNSISQIKCQLSIPVRRDAGEVRCF
13	Q61400	CEAMA_MOU	-----
14	Q2WEN9	CEA16_HUM	---PKDLVAVNRGFASERHLLSQPSGTIAGFAHTGREVCFPCNSLAVQKINITDTRITL
15	Q6UY09	CEA20_HUM	---KTTEDPSHETSQETPKKEHPTPESSSLSFPCNLSQVGRIRVEMOPDLEETVETKIPASRRGNSLSPVKKPPKFLMPP
16	Q7Z692	CEA19_HUM	-----

1060.....1070.....1080.....1090.....1100.....1110.....1120.....1130.....1140



```

1 P06731|CEAM5_HUM TSAGATVGMIGVLVGVALT-
2 P40199|CEAM6_HUM TSAWATVGTITIGVIARVALT-
3 P31997|CEAM8_HUM TSAHATVSGIMIGVIARVALT-
4 P13688|CEAM1_HUM TSASFSLTATELIYSEVKKQ-
5 Q14002|CEAM7_HUM TSAGTAVSIMGVLAGMALT-
6 Q16557|PSG3_HUMA MTKVVSAPSGTGHLPGINPT-
7 Q3KPI0|CEA21_HUM -----
8 P40198|CEAM3_HUM -----
9 P31809|CEAM1_MOU TSAPSSPRATEIIVYSEVKKK-
10 Q925P2|CEAM2_MOU TSAPSSPRATEIIVYSEVKKK-
11 P16573|CEAM1_RAT TSASSSP--TETVYSVVKKK-
12 Q63111|CEAM3_RAT VSNFVSSKTSLPVSLDYLIE-
13 Q61400|CEAMA_MOU -----
14 Q2WEN9|CEA16_HUM TIVTQGKTETEVEVLCVAPLG
15 Q6UY09|CEA20_HUM RLVSTVPRNMESIVVILGMQQ
16 Q7Z692|CEA19_HUM -----
.....1150.....1160.
    
```

可以看出，16 个序列内部存在一定的氨基酸残基位点，是哪一些具体的位点保守呢？这些位点的变异范围又如何呢？这些疑问又需要依赖下面的相关蛋白质的保守基序和进化树分析结果来予以说明。

## 分析之二：人癌胚抗原相关蛋白质的保守基序和进化树分析

### (1) MEME 分析：

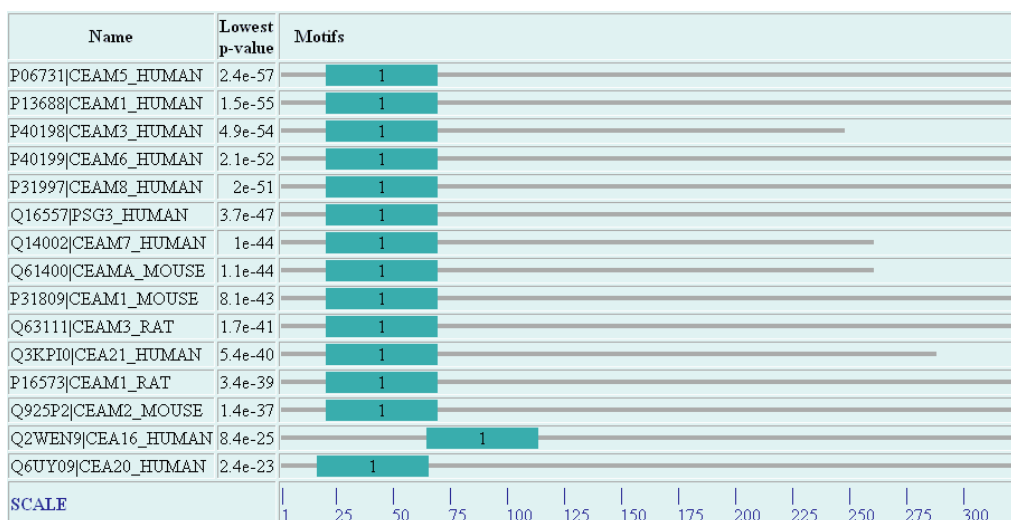
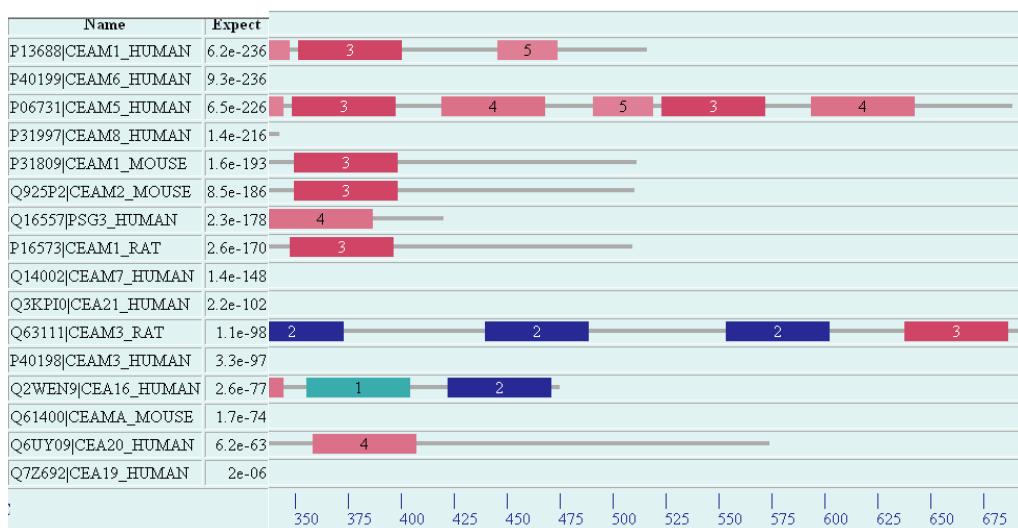
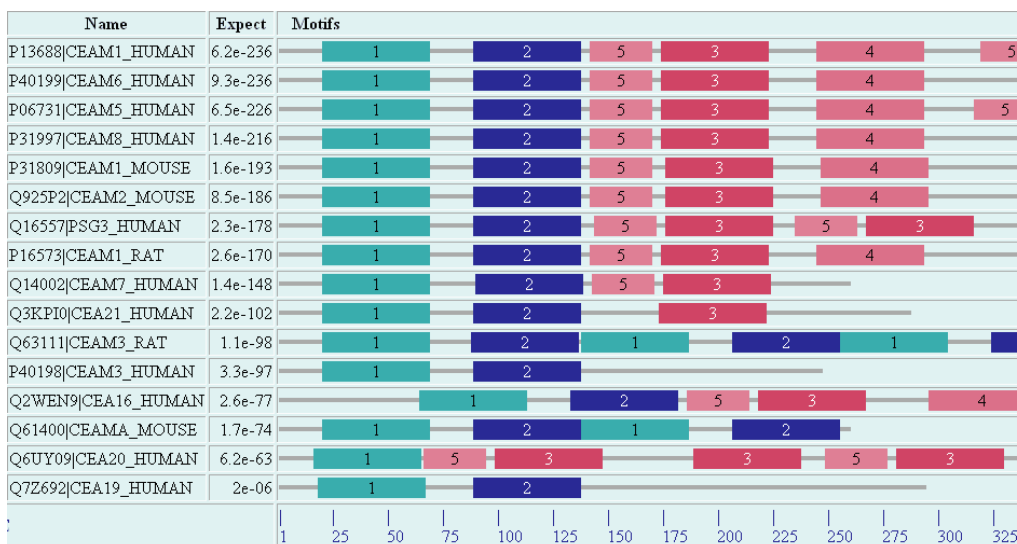
MOTIFS ./meme.html (peptide)

MOTIF WIDTH BEST POSSIBLE MATCH

```

-----
1 50 TASLLTFWNPPTTAQVTIEAMPFNVAEGKEVLLLHNLPHLFGYSWYKG
2 50 PGPAYSGRETIYPNGSLLFQNVTMNDTGFYTLHMIKRDFKNEEATGQFHV
3 50 WWFNGQSLPVSDRLQLSEGNRRTLTLFNVRRNDAGPYECEIWNPVSANRSD
4 50 HPGENLNSCHAASNPPAQYFWFINGKFQQSTQELFIPNITTNNSGSYMC
5 29 PKPSITSNNSNPVEDKDAVFTCEPETQN
    
```

Sequence Name	Description	E-value	Length
P13688 CEAM1_HUMAN	Carcinoembryonic antigen-...	6.2e-236	526
P40199 CEAM6_HUMAN	Carcinoembryonic antigen-...	9.3e-236	344
P06731 CEAM5_HUMAN	Carcinoembryonic antigen-...	6.5e-226	702
P31997 CEAM8_HUMAN	Carcinoembryonic antigen-...	1.4e-216	349
P31809 CEAM1_MOUSE	Carcinoembryonic antigen-...	1.6e-193	521
Q925P2 CEAM2_MOUSE	Carcinoembryonic antigen-...	8.5e-186	520
Q16557 PSG3_HUMAN	Pregnancy-specific beta-1...	2.3e-178	428
P16573 CEAM1_RAT	Carcinoembryonic antigen-...	2.6e-170	519
Q14002 CEAM7_HUMAN	Carcinoembryonic antigen-...	1.4e-148	265
Q3KPI0 CEA21_HUMAN	Carcinoembryonic antigen-...	2.2e-102	293
Q63111 CEAM3_RAT	Carcinoembryonic antigen-...	1.1e-98	709
P40198 CEAM3_HUMAN	Carcinoembryonic antigen-...	3.3e-97	252
Q2WEN9 CEA16_HUMAN	Carcinoembryonic antigen-...	2.6e-77	484
Q61400 CEAMA_MOUSE	Carcinoembryonic antigen-...	1.7e-74	265
Q6UY09 CEA20_HUMAN	Carcinoembryonic antigen-...	6.2e-63	585
Q7Z692 CEA19_HUMAN	Carcinoembryonic antigen-...	2e-06	300



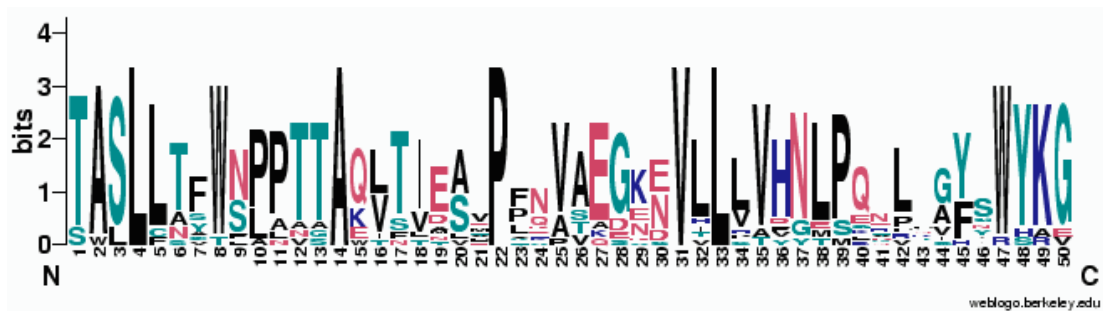
```

>P06731|CEAM5_HUMAN( start=21 )TASLLTFWNPPTAKLT IESTPFNVAEGKEVLLL VHNLPQHLFGYSWYKG
>P13688|CEAM1_HUMAN( start= 21 )TASLLTFWNPPTAQLT TESMPFNVAEGKEVLLL VHNLPQQLFGYSWYKG
>P40198|CEAM3_HUMAN( start= 21 )TASLLNFWNPPTAKLT IESMPLSVAEGKEVLLL VHNLPQHLFGYSWYKG
>P40199|CEAM6_HUMAN( start= 21 )TASLLTFWNPPTAKLT IESTPFNVAEGKEVLLL AHNLPQNRIGYSWYKG
>P31997|CEAM8_HUMAN( start= 21 )TASLFTFWNPPTAQLT IEAVPSNAAEGKEVLLL VHNLPQDPRGYNWYKG
>Q16557|PSG3_HUMAN( start= 21 )TALLNFWNLPTTAQVT IEAEPTKVS KGKDVLLL VHNLPQNLAGYIWKKG
    
```

```

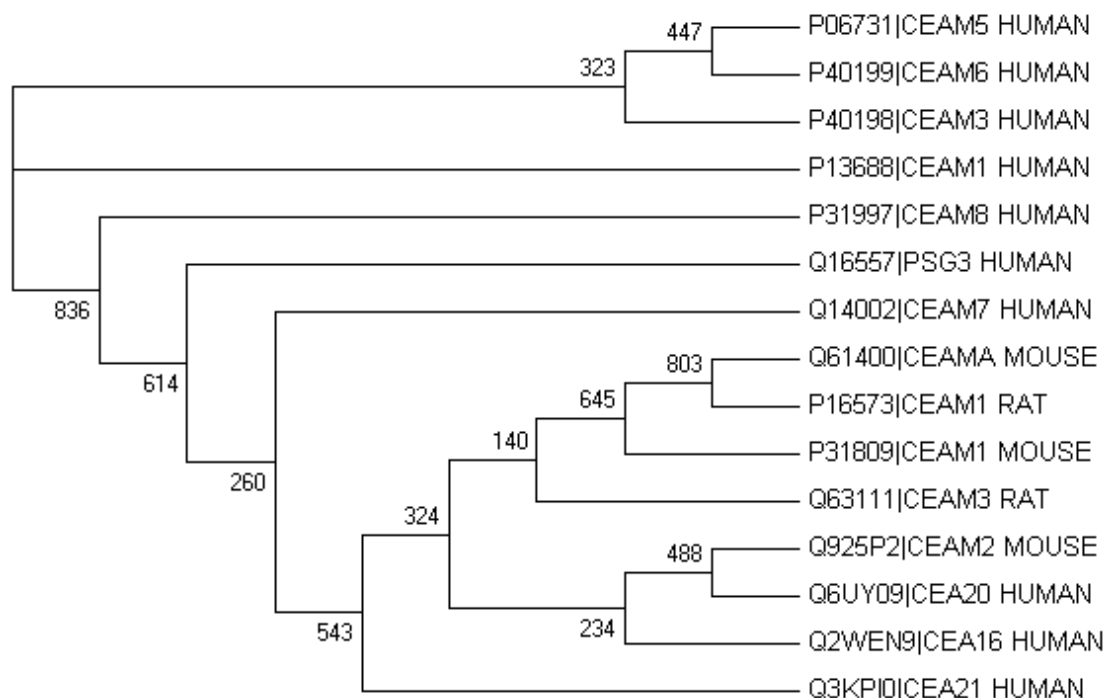
>Q14002|CEAM7_HUMAN( start= 21 )TASLLTFWNLPNSAQTNIDVVPFNVAEGKEVLLVHVNESQNLGYGNWYKG
>Q61400|CEAMA_MOUSE( start= 21 )TASLLTYWSPATTAQVTVEAVPPNVTADNNVLLLHVNLPQTLRVFYWYKG
>P31809|CEAM1_MOUSE( start= 21 )TASLLASWSPATTAQVTVEAVPPQVAEDNNVLLLHVNLPALGAFAWYKG
>Q63111|CEAM3_RAT( start= 21 )TASLLTCWLLPTTAQVSIKSLPPQVVEGENVLLHVDNLPENLIAFVWYKG
>Q3KPI0|CEA21_HUMAN( start= 21 )TASLLTFWNAPTAWLFIASAPFEVAEGENVHLSVVYLPENLYSYGWYKG
>P16573|CEAM1_RAT( start= 21 )TASLLTYWSPLTTAQTVDVAVPPNVVEESSVLLLTHNLPQEFQVFYWYKV
>Q925P2|CEAM2_MOUSE( start= 21 )TASLLASWSPPTTAQVTVMAPPLHAAEGNNVILVVYNNMKGVSAPSWHKG
>Q2WEN9|CEA16_HUMAN( start= 66 )SWLLSATFLNVGAEISITILEPAQPSEGDNVTLVHGLSGELLAYSWYAG
>Q6UY09|CEA20_HUMAN( start= 17 )SASLCTVWSPAAAQLTLNANPLDATQSEDVVLVFPVFGTPRTPQIHGRSRE
    
```

(2) Weblog 分析:



可以看出在这些蛋白质序列中存在着部分位点的氨基酸残基的不同程度的保守性，如上图中的 4 位的 P、14 位的 A、22 位的 P、31 位的 V 和 33 位的 L。这些保守氨基酸残基对于该类蛋白的功能表现的作用如何，有赖于从上述蛋白质的三维结构中得到些许证实，可惜，当前该类蛋白的结构研究很少。

(3) 进化树分析:

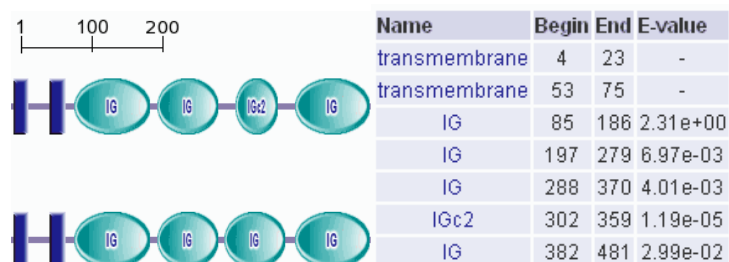


由此对基序 1 进行的进化树分析图可以初步得出上述几种蛋白之间的进化关系，结果与前面进行的序列比对结果差异较大，不仅说明比对与进化树构建两种分析途径的不同，也说明 CEA 家族成员内部之间的明显差异性，研究 CEA 家族面临的困难可能要超过以往。

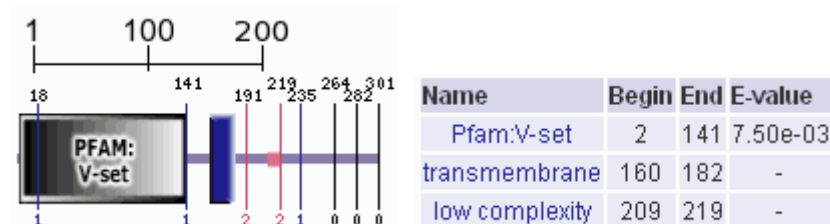
### 分析之三：人癌胚抗原相关蛋白质的结构域和跨膜分析

#### (1) SMART 分析:

>Q2WEN9|CEA16\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 16 - Homo sapiens (Human).

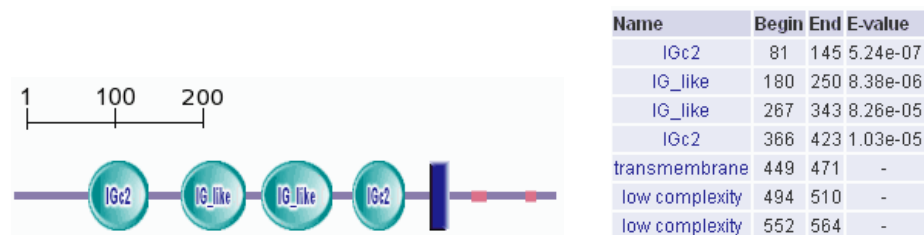


>Q7Z692|CEA19\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 19 - Homo sapiens (Human).

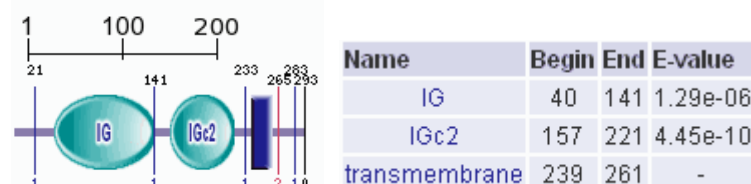


#### Immunoglobulin V-set domain

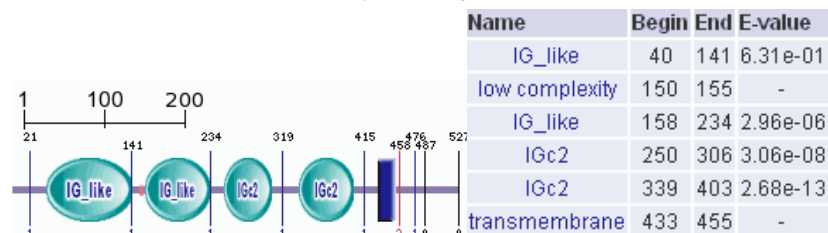
>Q6UY09|CEA20\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 20 - Homo sapiens (Human).



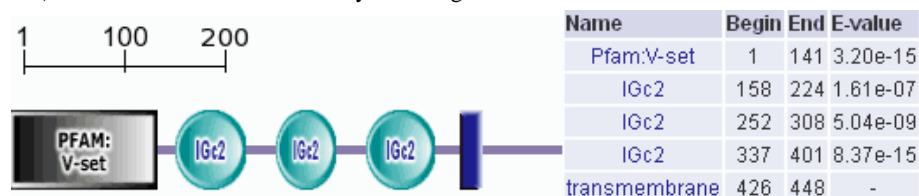
>Q3KPI0|CEA21\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 21 - Homo sapiens (Human).



>P13688|CEAM1\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 1 - Homo sapiens (Human).



>P31809|CEAM1\_MOUSE Carcinoembryonic antigen-related cell adhesion molecule 1 - Mus musculus (Mouse).



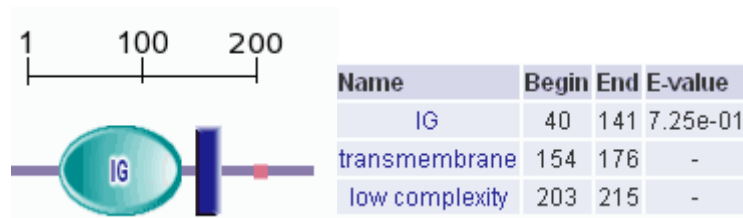
>P16573|CEAM1\_RAT Carcinoembryonic antigen-related cell adhesion molecule 1 - Rattus norvegicus (Rat)



>Q925P2|CEAM2\_MOUSE Carcinoembryonic antigen-related cell adhesion molecule 2 - Mus musculus (Mouse).



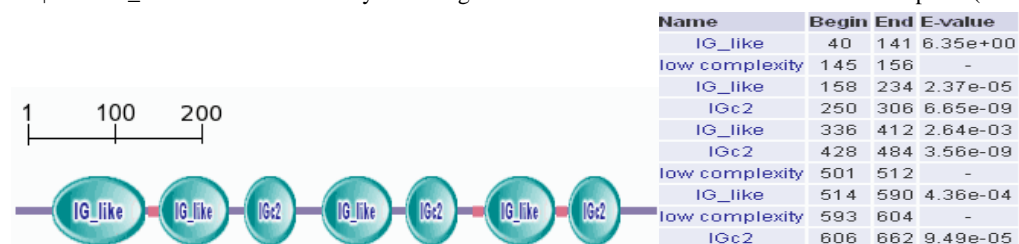
>P40198|CEAM3\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 3 - Homo sapiens (Human).



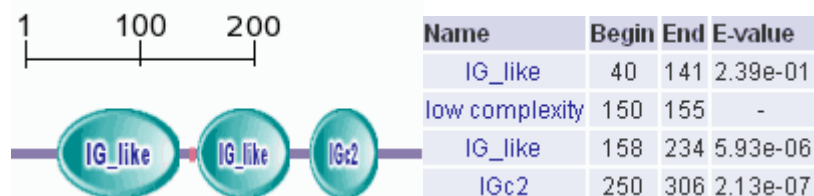
>Q63111|CEAM3\_RAT Carcinoembryonic antigen-related cell adhesion molecule 3 - Rattus norvegicus (Rat).



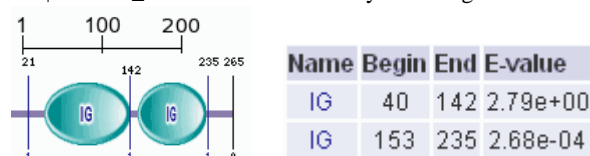
>P06731|CEAM5\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 5 - Homo sapiens (Human).



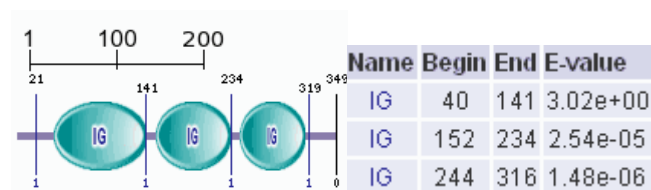
>P40199|CEAM6\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 6 - Homo sapiens (Human).



>Q14002|CEAM7\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 7 - Homo sapiens (Human).



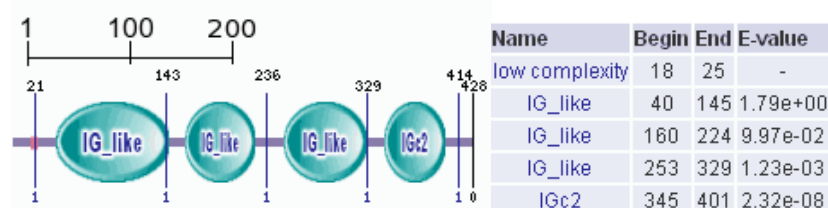
>P31997|CEAM8\_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 8 - Homo sapiens (Human).



>Q61400|CEAMA\_MOUSE Carcinoembryonic antigen-related cell adhesion molecule 10 - Mus musculus (Mouse).

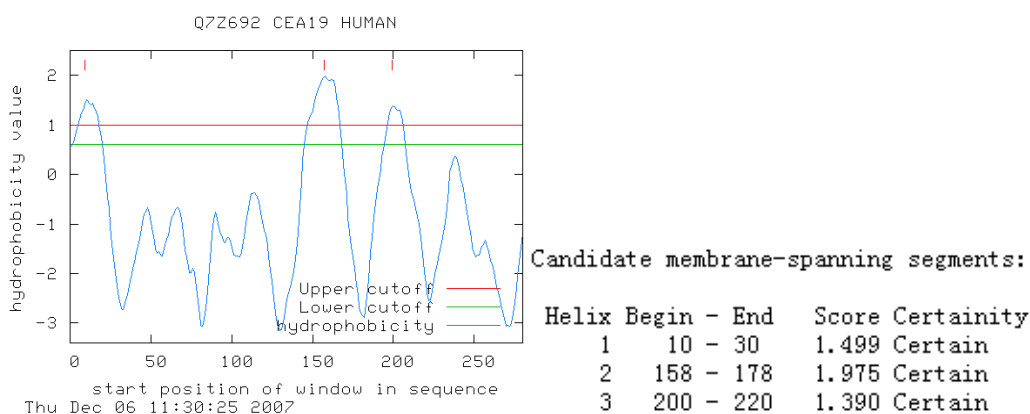
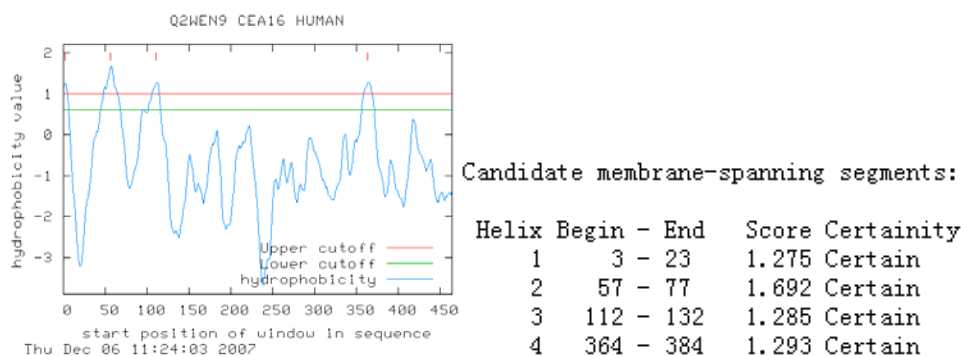


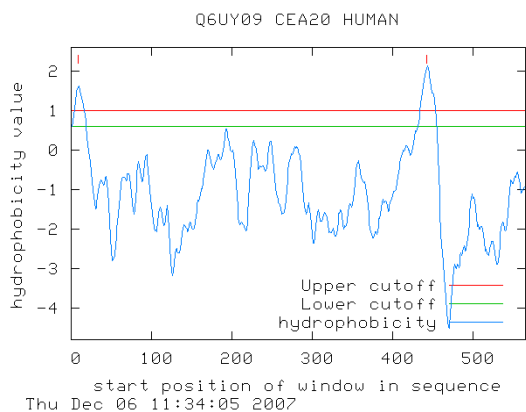
>Q16557|PSG3\_HUMAN Pregnancy-specific beta-1-glycoprotein 3 - Homo sapiens (Human).



SMART 分析可以初步解释在前面序列比对中出现的重复点阵区应该是免疫球蛋白结构域区，也可以一定程度上解释出 CEA5 与 CEA3 的综合相似程度高，还可以看出 CEA 家族成员普遍具有免疫球蛋白结构域，从这一点出发，可以利用已知的免疫球蛋白三维结构来推测未知结构的 CEA 家族蛋白。

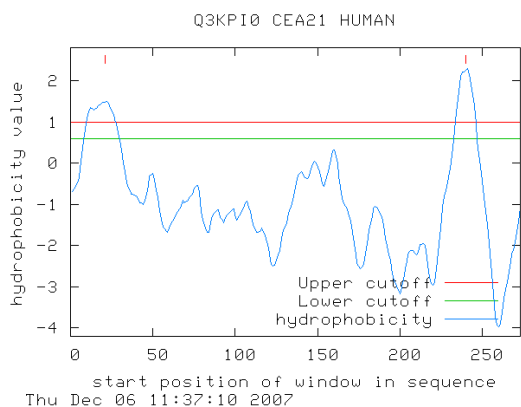
## (2) TopPred 分析:





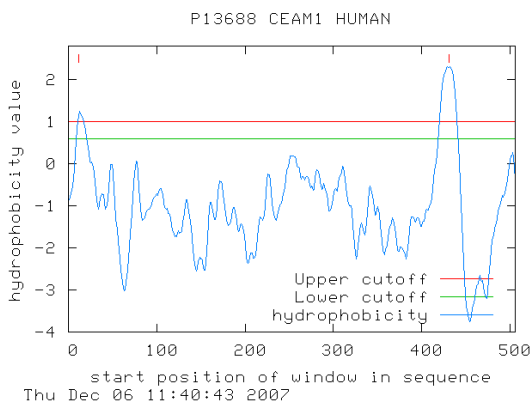
Candidate membrane-spanning segments:

Helix	Begin - End	Score	Certainty
1	10 - 30	1.618	Certain
2	443 - 463	2.134	Certain



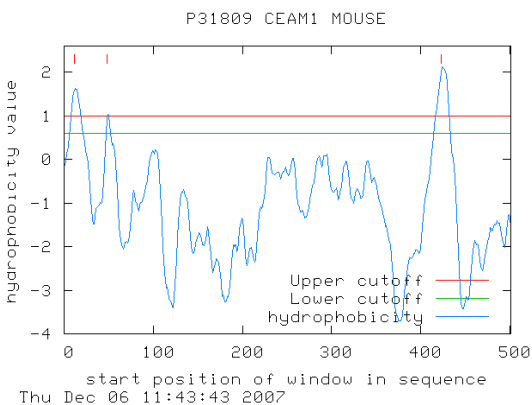
Candidate membrane-spanning segments:

Helix	Begin - End	Score	Certainty
1	22 - 42	1.486	Certain
2	241 - 261	2.285	Certain



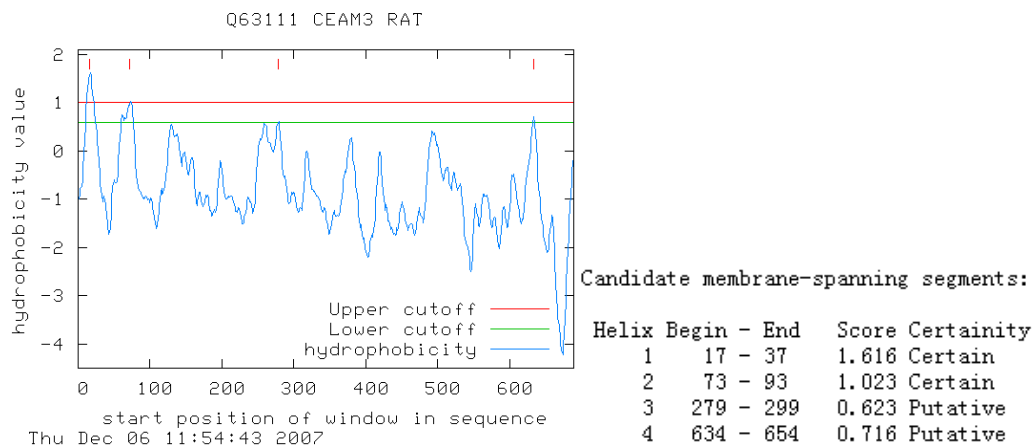
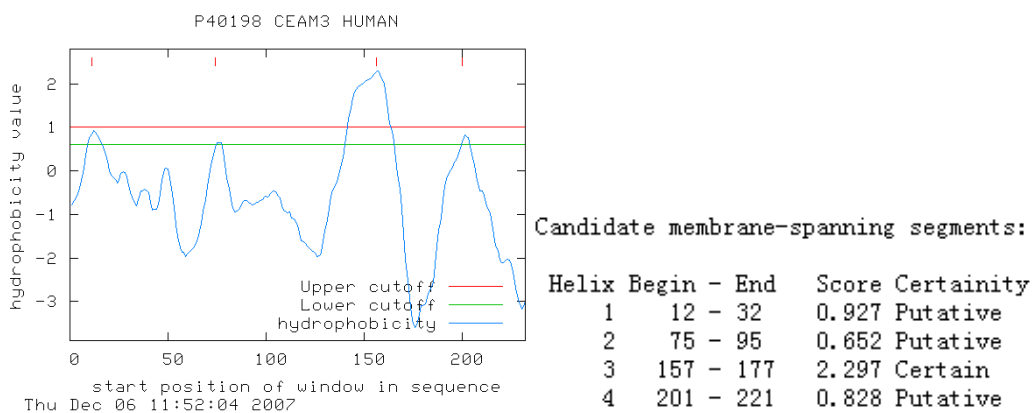
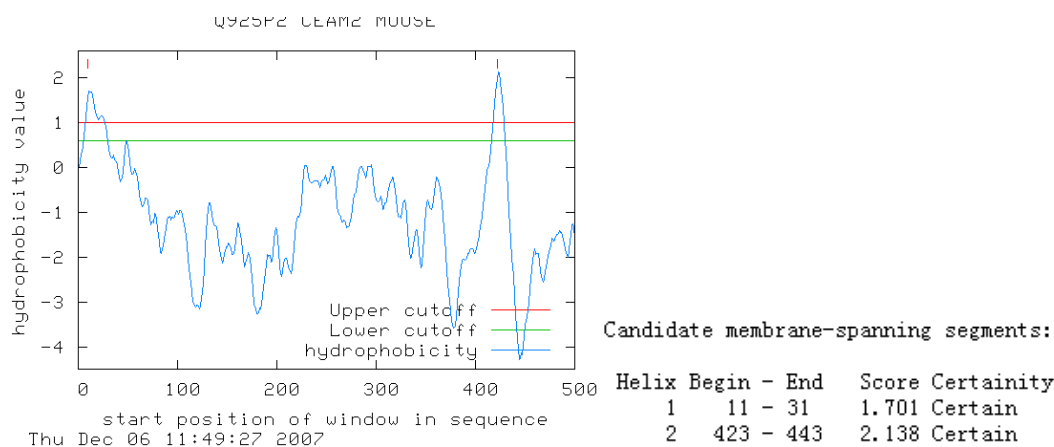
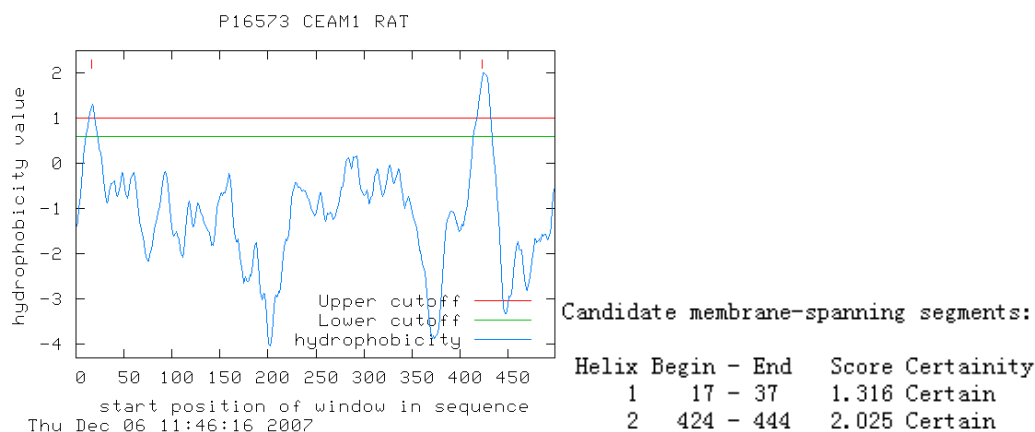
Candidate membrane-spanning segments:

Helix	Begin - End	Score	Certainty
1	13 - 33	1.249	Certain
2	432 - 452	2.313	Certain



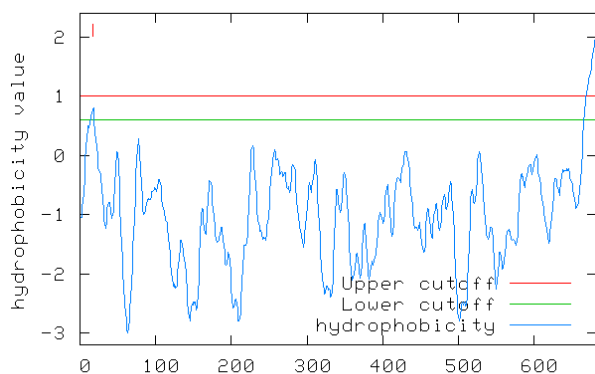
Candidate membrane-spanning segments:

Helix	Begin - End	Score	Certainty
1	13 - 33	1.630	Certain
2	49 - 69	1.031	Certain
3	424 - 444	2.131	Certain





P06731 CEAM5 HUMAN

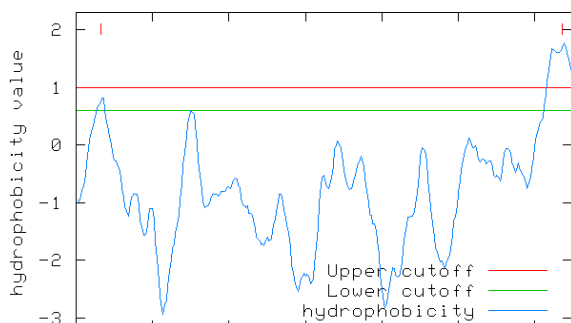


start position of window in sequence  
Thu Dec 06 11:56:49 2007

Candidate membrane-spanning segments:

Helix	Begin	End	Score	Certainty
1	18	38	0.807	Putative
2	682	702	2.070	Certain

P40199 CEAM6 HUMAN

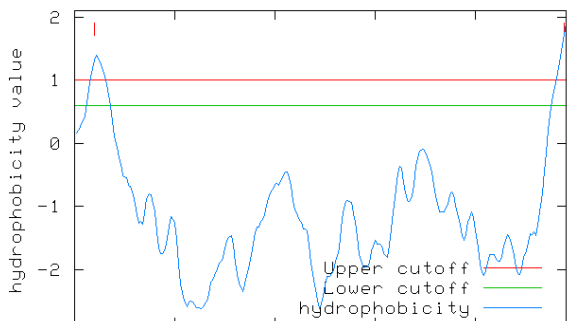


start position of window in sequence  
Thu Dec 06 12:00:12 2007

Candidate membrane-spanning segments:

Helix	Begin	End	Score	Certainty
1	17	37	0.823	Putative
2	319	339	1.768	Certain

Q14002 CEAM7 HUMAN

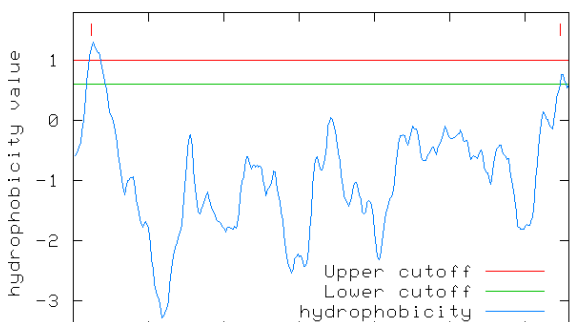


start position of window in sequence  
Thu Dec 06 12:02:18 2007

Candidate membrane-spanning segments:

Helix	Begin	End	Score	Certainty
1	11	31	1.403	Certain
2	245	265	1.909	Certain

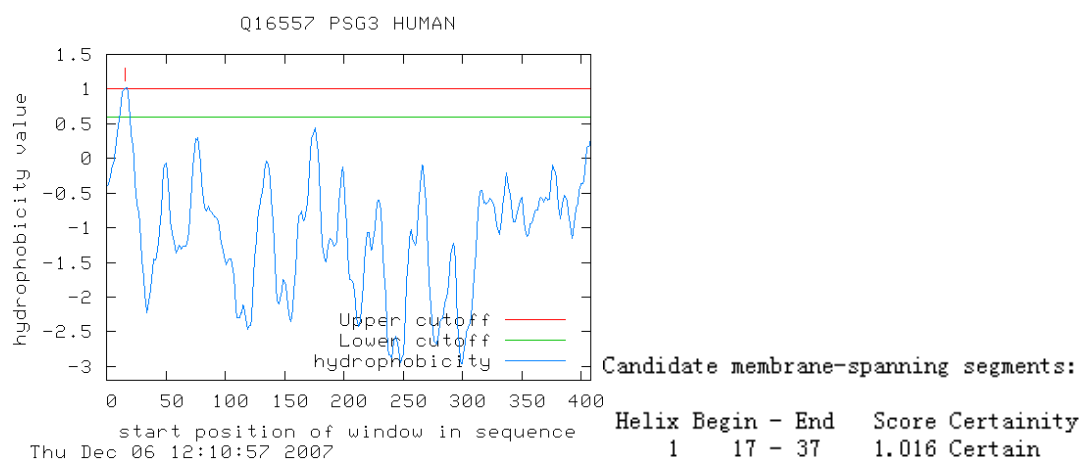
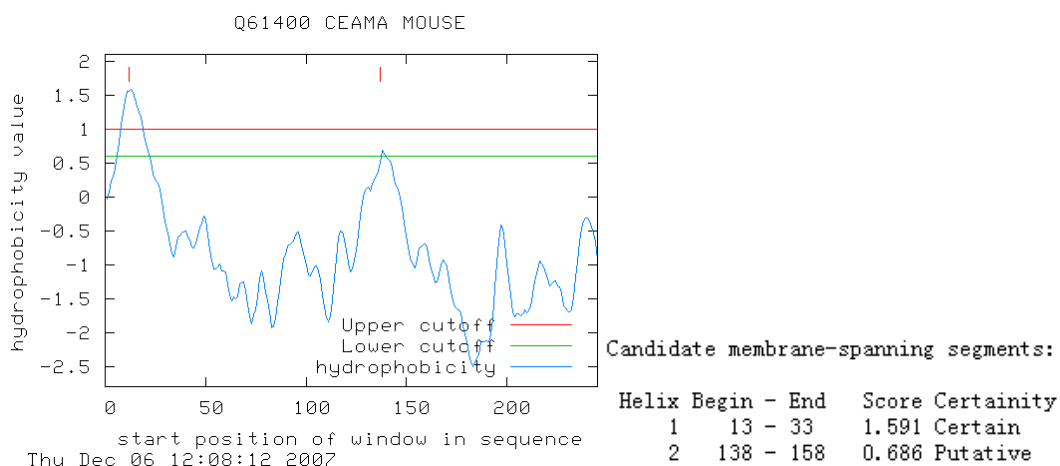
P31997 CEAM8 HUMAN



start position of window in sequence  
Thu Dec 06 12:04:57 2007

Candidate membrane-spanning segments:

Helix	Begin	End	Score	Certainty
1	13	33	1.310	Certain
2	324	344	0.758	Putative



结果表明上述蛋白质都存在非常明显的跨膜区域,这些特点对于研究相关抗原蛋白的详细功能应该有用。下一步,可以考虑参照免疫球蛋白结构来预测人癌胚抗原蛋白的三维结构,具体分析还有待于本小组成员的进一步学习。