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

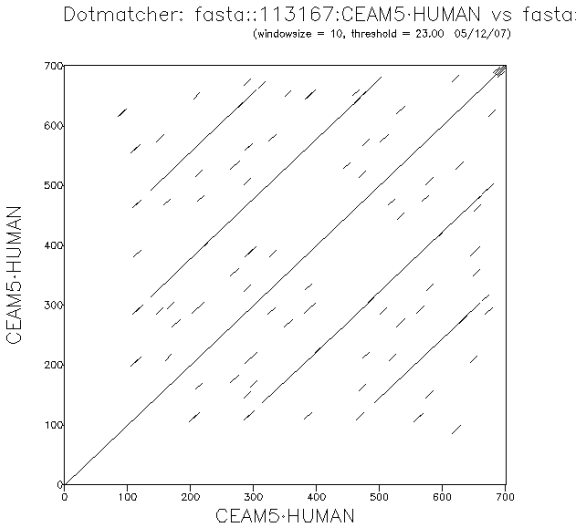
人癌胚抗原是一种富含多糖的蛋白复合物，主要存在于直、结肠癌组织和胎儿肠粘膜内，属性为膜表面蛋白质。在个体发育过程中，人癌胚抗原蛋白只在胎儿时期大量表达，随着发育过程的不断深入，许多人癌胚抗原的表达明显减少甚至停滞表达。成年细胞发生癌变后会出现去分化现象，导致原先被关闭表达的人癌胚抗原再次活跃起来，重新达到胎儿时期的表达状态。基于其生理功能的重要性，研究人癌胚抗原应该具有很强烈的现实意义。

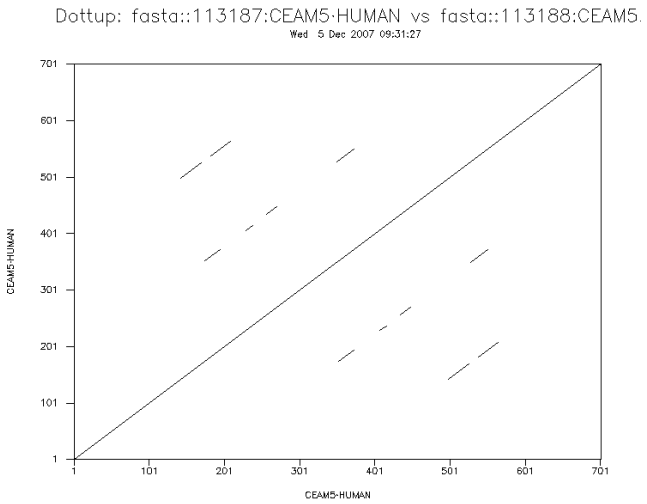
分析之一：人癌胚抗原相关蛋白质的序列搜索与比对

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：

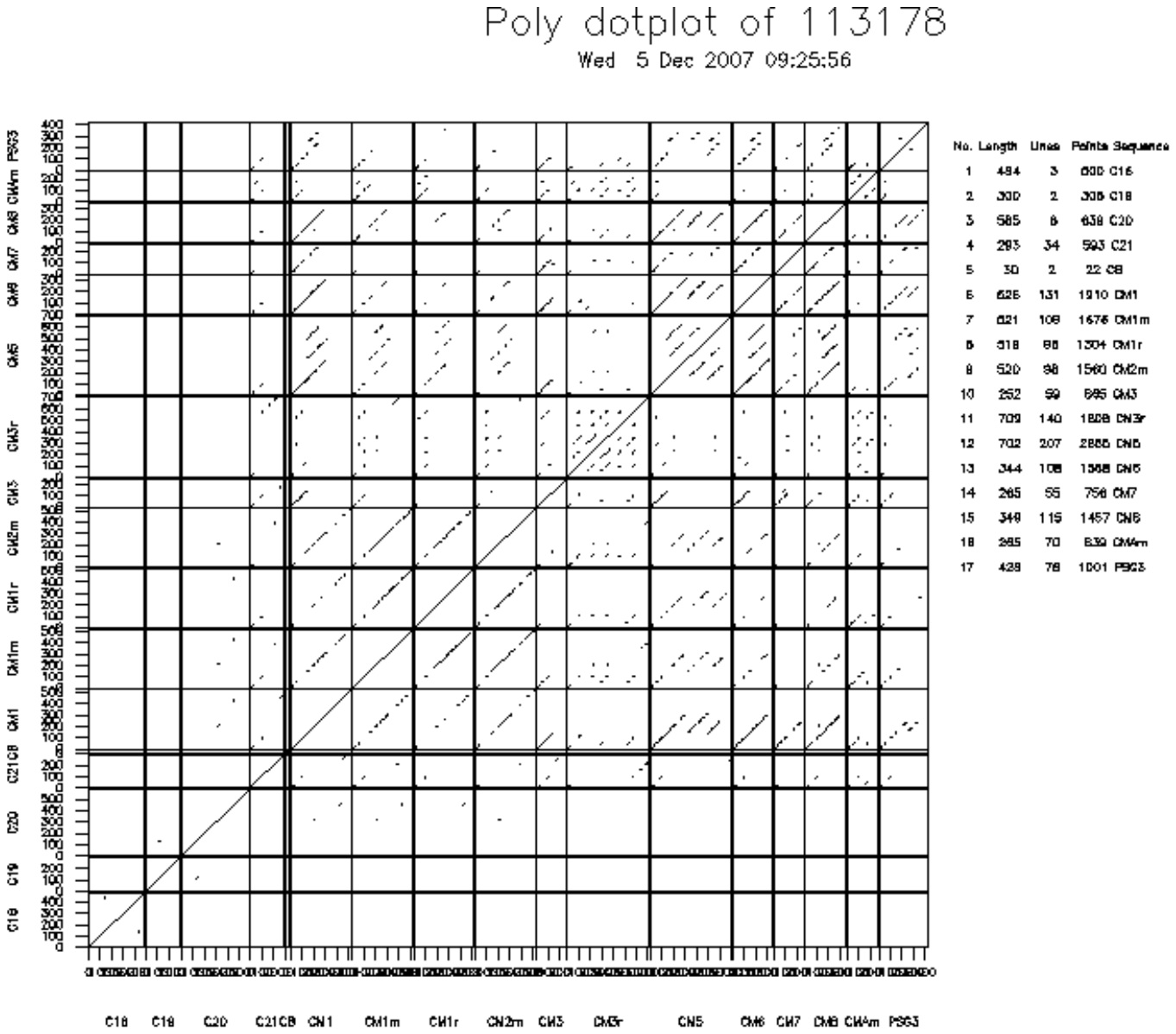
（1）CEAM5_HUMAN 自身序列的 Dotmatcher 和 Dottup 比对





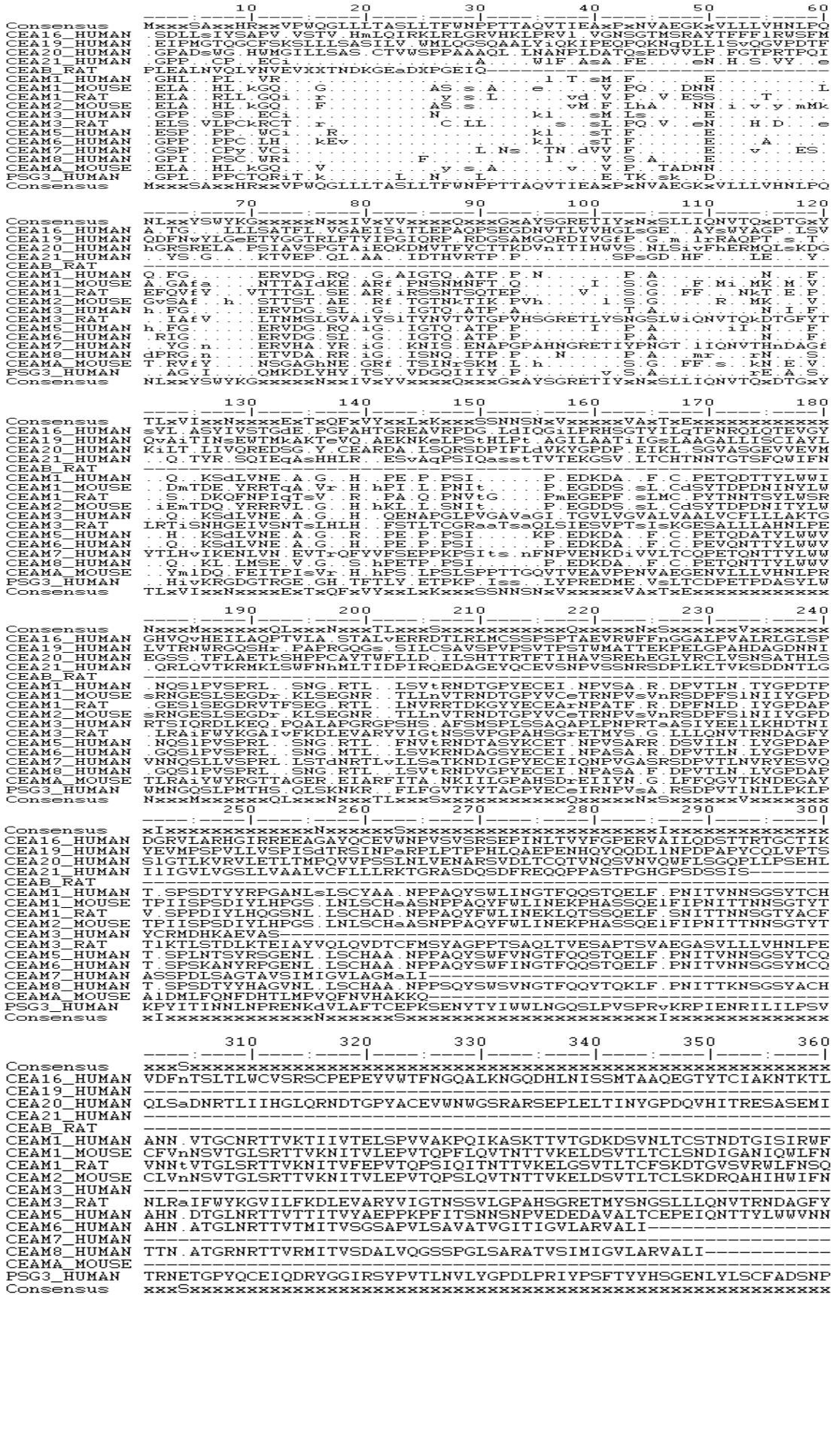
可以看出：CEAM5_HUMAN 自身序列内部存在不同程度的序列重复，这些重复序列与后面将要研究的免疫球蛋白结构域又有如何的关联呢？需要接下来的进一步工具分析。

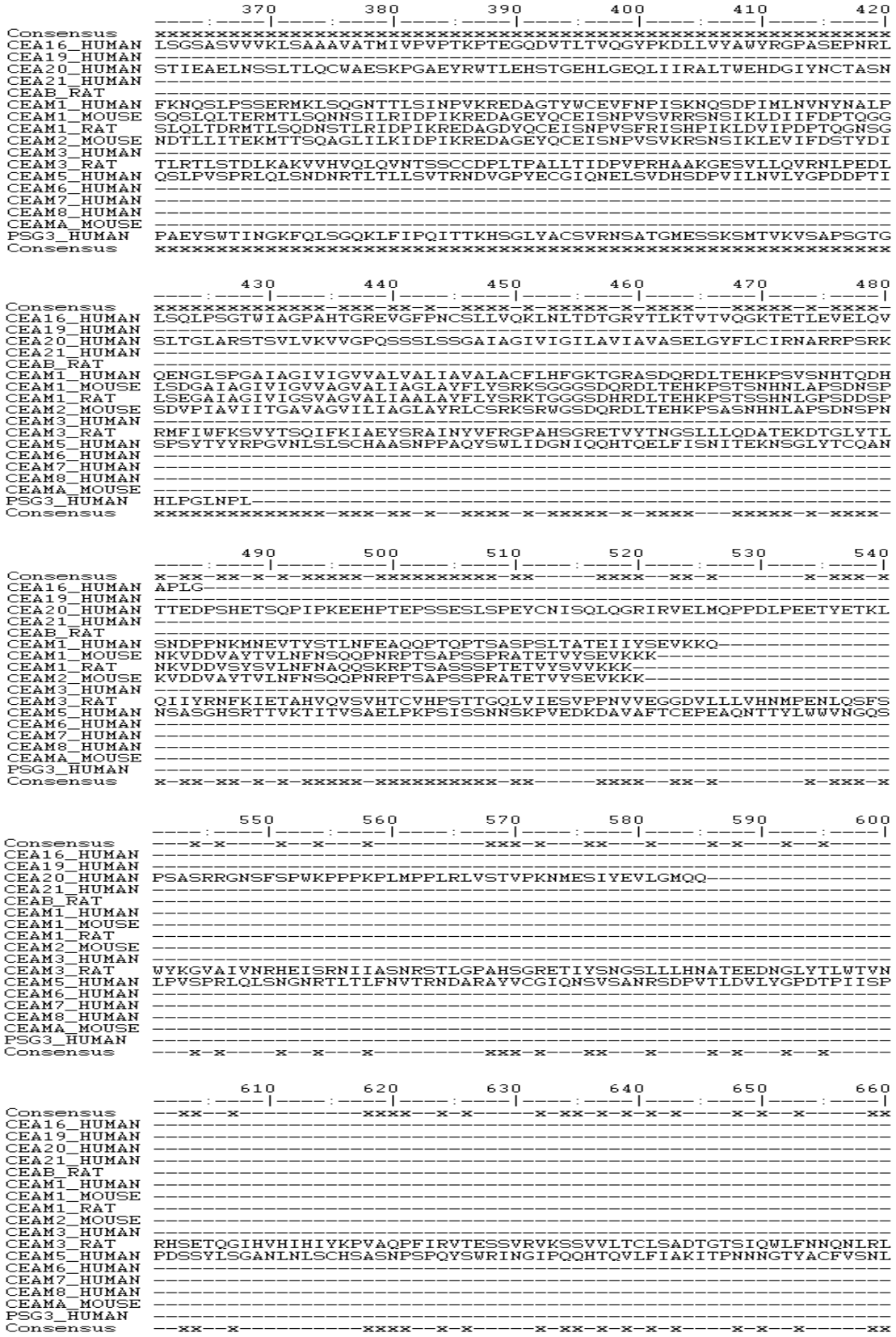
（2）17 条蛋白质信息的 Poly dotplot 比对：

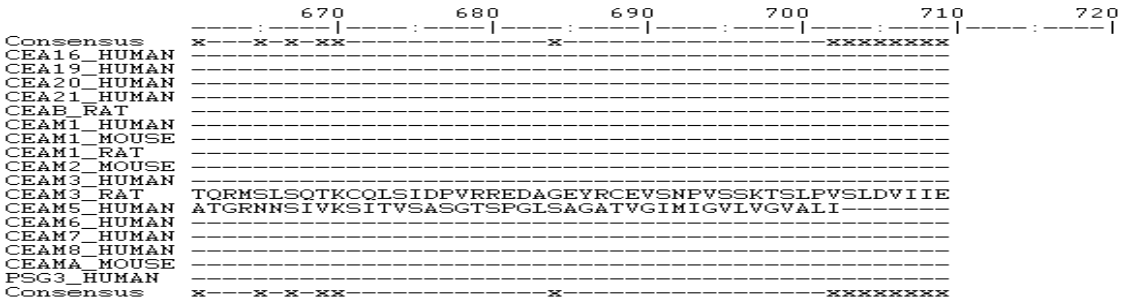


可以看出：CEA16_HUMAN (Q2WEN9)、CEA19_HUMAN (Q7Z692)和 CEA20_HUMAN (Q6UY09)、三个序列之间以及与其它 14 个序列之间的序列差异很大，Poly dotplot 比对结果中几乎找不到重叠片段。发现序列 CEAM3_RAT 自身比对结果中有众多杂点区，显示出该序列存在许多小片段的重复。还发现序列 CEAM5_HUMAN 与 CEAM1_HUMAN、CEAM6_HUMAN 、CEAM8_HUMAN 三个序列间存在较大片段的重复。为了看一看它们相互之间的详细序列位点差异，我们有必要进行以下初步蛋白质信息分析和 Clustalw 全序列信息比对分析。

（3）Showalign 和 emma 分析 17 条蛋白质信息：




370 380 390 400 410 420
Consensus
CEA16_HUMAN LSGSASVVVKLSAAAVATMIVPVPTKPTEGQDVTLTVQGVPKDLLVYAWYRGPASEPNRL
CEA19_HUMAN
CEA20_HUMAN STIEAELNSSLTLQCWAESKPGAEYRWTLEHSTGEHLGEQLIIRALTWEHDGIYNCTASN
CEA21_HUMAN
CEAB_RAT
CEAM1_HUMAN FKNQSLPSSERMKLSQGNTTLSINPVKREDAGTYWCEVFNPISKNQSDPIMLNVNYNALP
CEAM1_MOUSE SQSLQLTERMTLSQNNSILRIDPIKREDAGEYQCEISNPVSVRRSNSIKLDIIFDPTQGG
CEAM1_RAT SLQLTDRMTLSQDNSTLRIDPIKREDAGDYQCEISNPVSFRISHPIKLDVIPDPTQGNSG
CEAM2_MOUSE NDTLLITEKMTTSQAGLILKIDPIKREDAGEYQCEISNPVSVKRSNSIKLEVIFDSTYDI
CEAM3_HUMAN
CEAM3_RAT TLRTLSTDLKAKVVHVQLQVNTSSCCDPLTPALLTIDPVPRHAAKGESVLLQVRNLPEDL
CEAM5_HUMAN QSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTI
CEAM6_HUMAN
CEAM7_HUMAN
CEAM8_HUMAN
CEAMA_MOUSE
PSG3_HUMAN PAEYSWTINGKFQLSGQKLFIPQITTKHSGLYACSVRNSATGMESSKSMTVKVSAPSGTG
Consensus
430 440 450 460 470 480
Consensus
CEA16_HUMAN LSQLPSGTWIAGPAHTGREVGFPNCSLLVQKLNLTDTGRYTLKTVTVQGKTETLEVELQV
CEA19_HUMAN
CEA20_HUMAN SLTGLARSTSVLVKVVGPQSSSLSSGAIAGIVIGILAVIAVASELGYFLCIRNARRPSRK
CEA21_HUMAN
CEAB_RAT
CEAM1_HUMAN QENGLSPGAIAGIVIGVVALVALIAVALACFLHFGKTGRASDQRDLTEHKPSVSNHTQDH
CEAM1_MOUSE LSDGAIAGIVIGVVAGVALIAGLAYFLYSRKSGGGSDQRDLTEHKPSTSNHNLAPSDNSP
CEAM1_RAT LSEGAIAGIVIGSVAGVALIAALAYFLYSRKTGGGSDHRDLTEHKPSTSSHNLGPSDDSP
CEAM2_MOUSE SDVPIAVIITGAVAGVILIAGLAYRLCSRKSRWGSDQRDLTEHKPSASNHNLAPSDNSPN
CEAM3_HUMAN
CEAM3_RAT RMFIWFKSVYTSQIFKIAEYSRAINYVFRGPAHSGRETVYTNGSLLLQDATEKDTGLYTL
CEAM5_HUMAN SPSYTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQAN
CEAM6_HUMAN
CEAM7_HUMAN
CEAM8_HUMAN
CEAMA_MOUSE
PSG3_HUMAN HLPGLNPL
Consensus
490 500 510 520 530 540
Consensus
CEA16_HUMAN APLG
CEA19_HUMAN
CEA20_HUMAN TTEDPSHETSQPIPKEEHPTEPSSESLSPEYCNISQLQGRIRVELMQPPDLPEETYETKL
CEA21_HUMAN
CEAB_RAT
CEAM1_HUMAN SNDPPNKMNEVTYSTLNFEAQQPTQPTSASPSLTATEIIYSEVKKQ
CEAM1_MOUSE NKVDDVAYTVLNFNSQQPNRPTSAPSSPRATETVYSEVKKK
CEAM1_RAT NKVDDVSYSVLNFNAQQSKRPTSASSSPTETVYSVVKKK
CEAM2_MOUSE KVDDVAYTVLNFNSQQPNRPTSAPSSPRATETVYSEVKKK
CEAM3_HUMAN
CEAM3_RAT QIIYRNFKIETAHVQVSVHTCVHPSTTGQLVIESVPPNVVEGGDVLLLVHNMPENLQSFS
CEAM5_HUMAN NSASGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQS
CEAM6_HUMAN
CEAM7_HUMAN
CEAM8_HUMAN
CEAMA_MOUSE
PSG3_HUMAN
Consensus
550 560 570 580 590 600
Consensus
CEA16_HUMAN
CEA19_HUMAN
CEA20_HUMAN PSASRRGNSFSPWKPPPKPLMPPLRLVSTVPKNMESIYEVLGMQQ
CEA21_HUMAN
CEAB_RAT
CEAM1_HUMAN
CEAM1_MOUSE
CEAM1_RAT
CEAM2_MOUSE
CEAM3_HUMAN
CEAM3_RAT WYKGVAIVNRHEISRNIIASNRSTLGPAHSGRETIVSNGSLLLHNATEEDNGLYTLWTVN
CEAM5_HUMAN LPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISP
CEAM6_HUMAN
CEAM7_HUMAN
CEAM8_HUMAN
CEAMA_MOUSE
PSG3_HUMAN
Consensus
610 620 630 640 650 660
Consensus
CEA16_HUMAN
CEA19_HUMAN
CEA20_HUMAN
CEA21_HUMAN
CEAB_RAT
CEAM1_HUMAN
CEAM1_MOUSE
CEAM1_RAT
CEAM2_MOUSE
CEAM3_HUMAN
CEAM3_RAT RHSETQGIHVHIHIYKPVAQPFIRVTESSVRVKSSVVLTCLSADTGTSIQWLFNNQNLRL
CEAM5_HUMAN PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNL
CEAM6_HUMAN
CEAM7_HUMAN
CEAM8_HUMAN
CEAMA_MOUSE
PSG3_HUMAN
Consensus



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

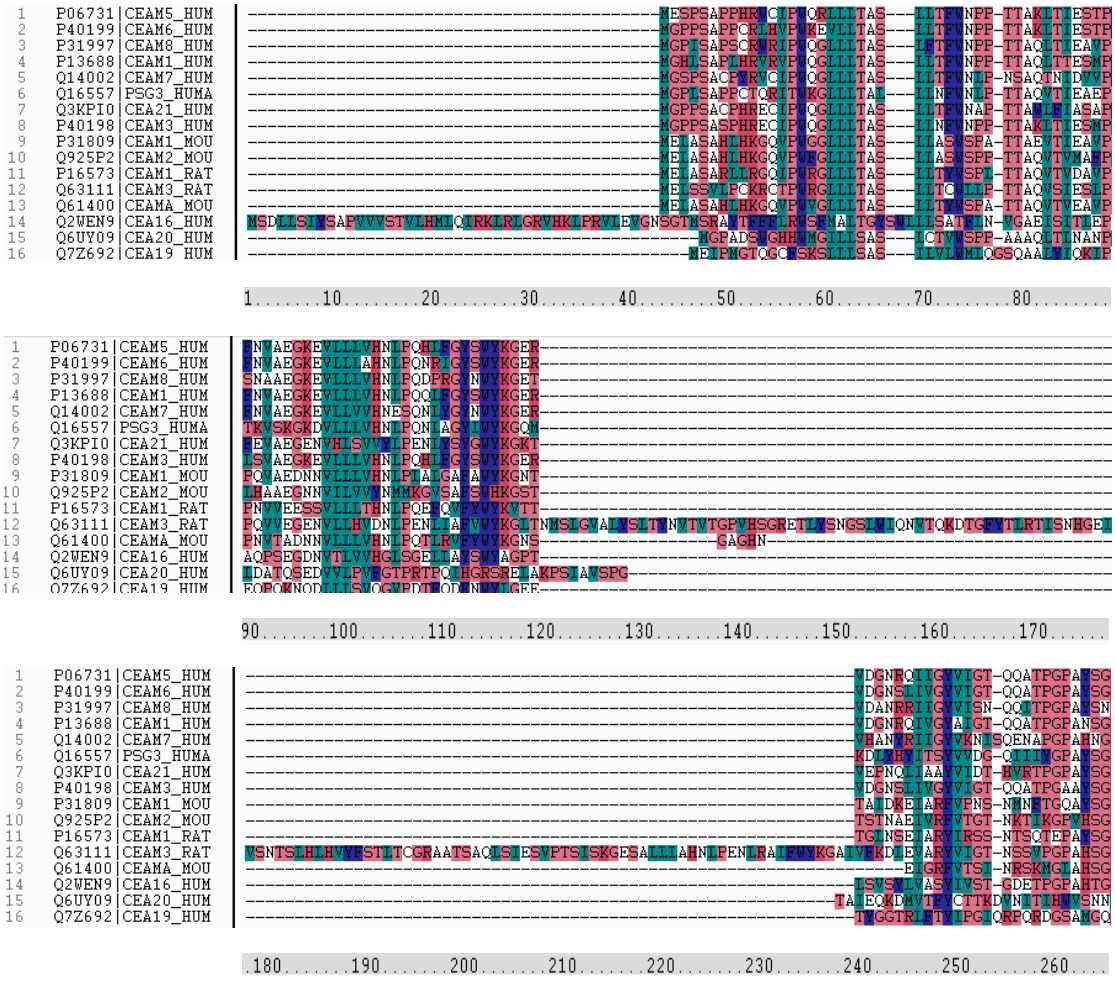
（4）Needle 和 Water 分析 17 条蛋白质信息：

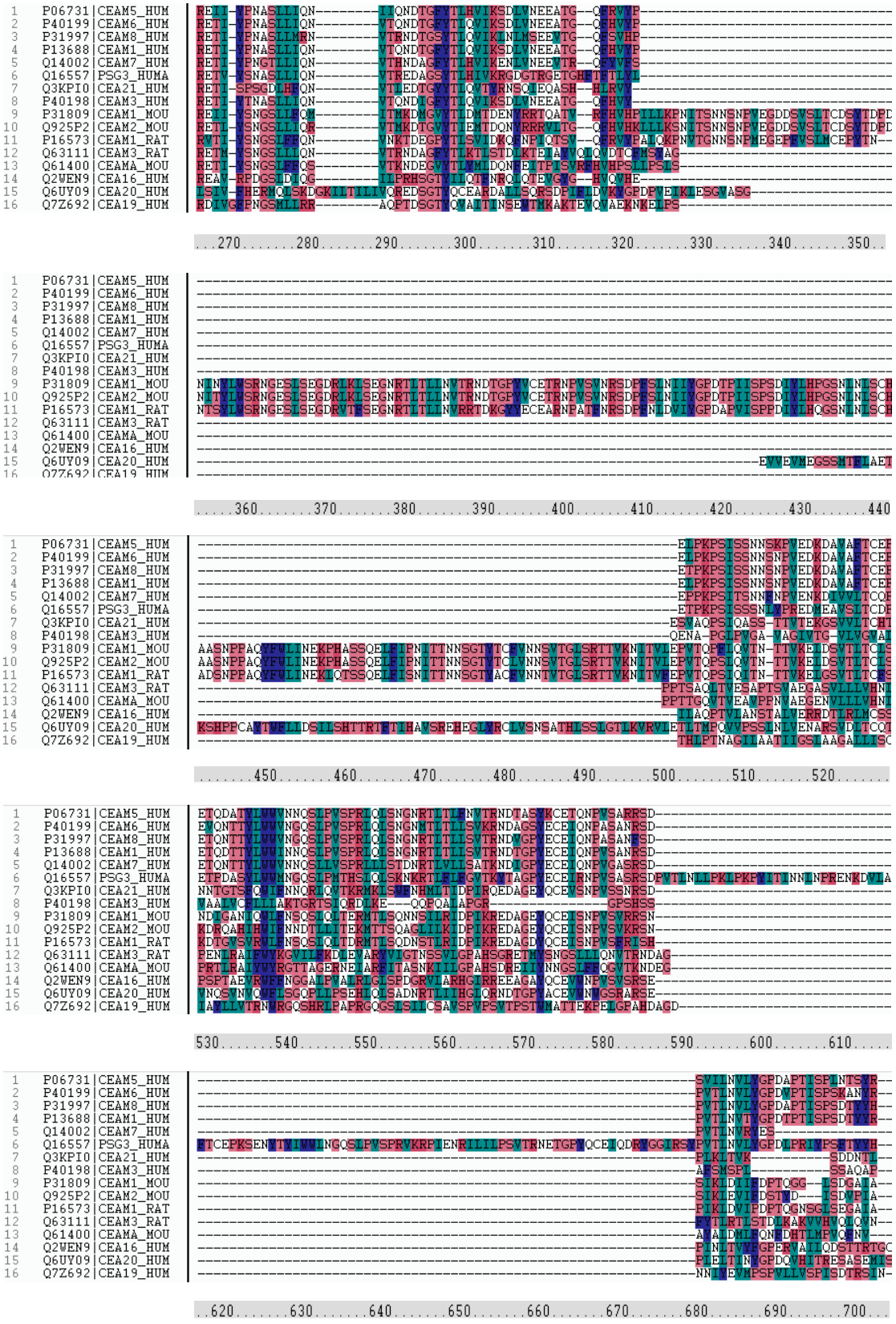
Alignment pairs	Alignment types	length	scope	identity	similarity	gaps
CEAM5＆CEA16	**needle**	772	450.5	18.5%	28.1%	46.4%
	water	401	474.5	31.4%	46.9%	8.7%
CEAM5＆CEA19	**needle**	723	148.5	11.8%	19.6%	61.4%
	water	395	157	20.3%	33.9%	34.7%
CEAM5＆CEA20	**needle**	798	587	21.6%	33.6%	38.7%
	water	521	601	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%

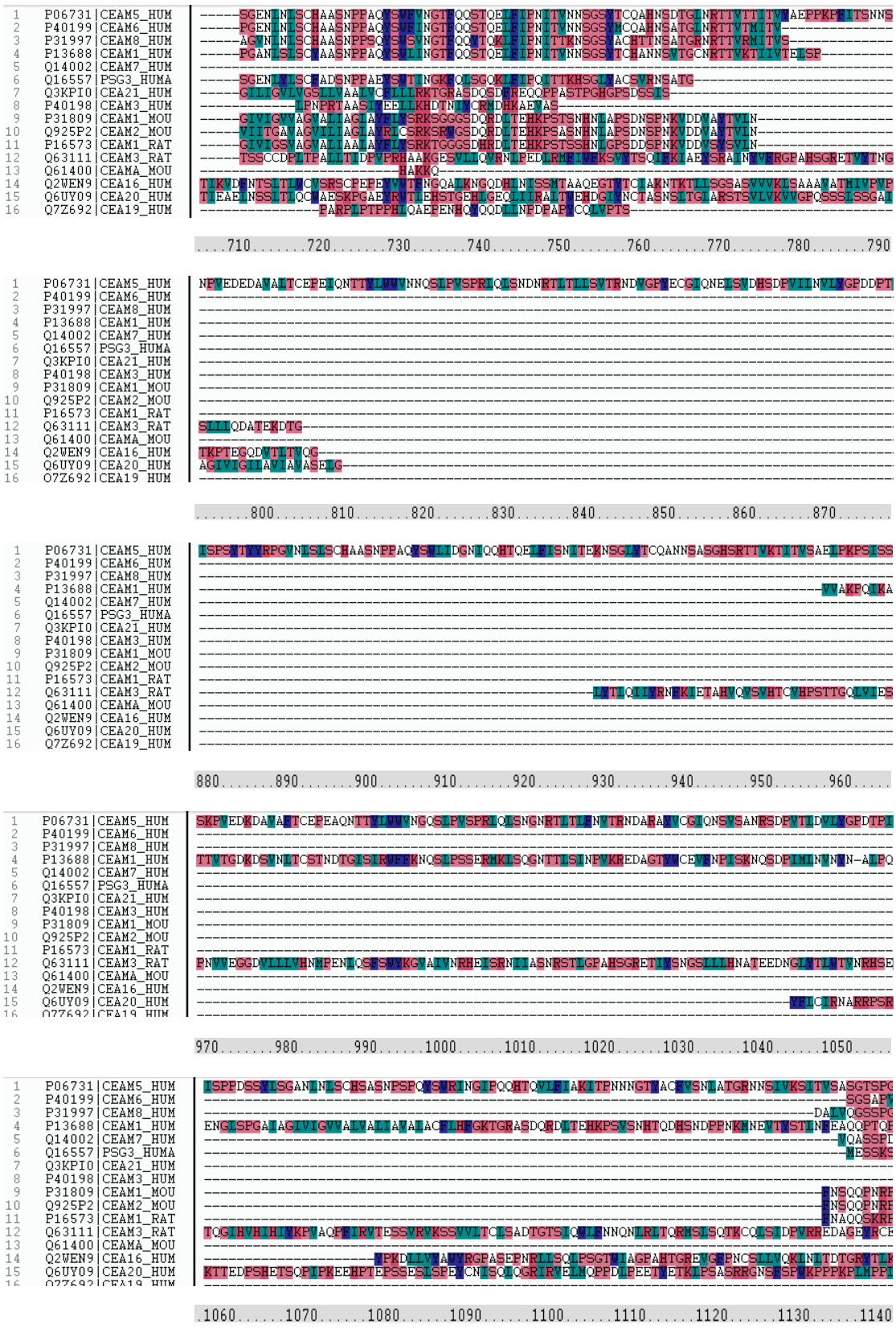
CEAM5＆CEAM8	**needle**	707	1255.5	34.9%	38.3%	51.3%
	water	314	1272	76.8%	82.8%	0%
CEAM5＆CEAMA-MO	**needle**	704	343.5	14.6%	20.3%	62.6%
	water	297	352	33%	45.8%	19.2%
CEAM5＆CEAPSG	**needle**	711	1135.5	33.6%	40.8%	41.1%
	water	496	1143	47.4%	57.7%	17.5%

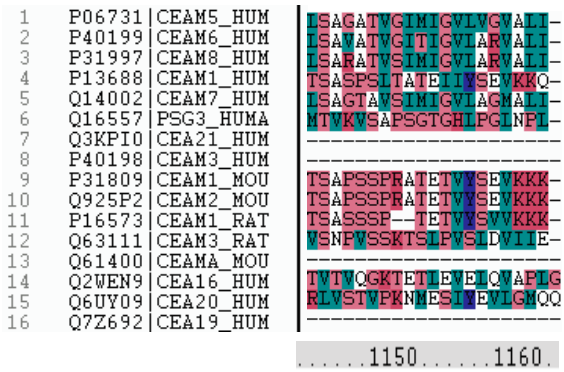
可以看出，就全局比对而言，CEAM1_HUMAN 与 CEAM5_HUMAN 的序列相似性最高，为 52.2%；就局部比对而言，CEAM1_HUMAN、CEAM3_HUMAN、CEAM6_HUMAN、CEAM7_HUMAN、CEAM8_HUMAN 都与 CEAM5_HUMAN 存在较高的序列相似性。就与 CEAM5_HUMAN 的整体比对效果而言，CEAM1_HUMAN 综合成绩应该是最好的。同时我们还可以发现，尽管人和大、小鼠的形态差异较大，但是 CEAM1_MOUSE、CEAM1_RAT 和 CEAM2_MOUSE 三个蛋白质序列与人的 CEAM5_HUMAN 也存在较高的序列相似性。

（4）Clustalw 全序列分析 16 条蛋白质（除去了其中的 CEAB_RAT）信息：




P06731|CEAM5_HUM
P40199|CEAM6_HUM
P31997|CEAM8_HUM
P13688|CEAM1_HUM
Q14002|CEAM7_HUM
Q16557|PSG3_HUMA
Q3KPI0|CEA21_HUM
P40198|CEAM3_HUM
P31809|CEAM1_MOU
Q925P2|CEAM2_MOU
P16573|CEAM1_RAT
Q63111|CEAM3_RAT
Q61400|CEAMA_MOU
Q2WEN9|CEA16_HUM
Q6UY09|CEA20_HUM
Q7Z692|CEA19_HUM


P06731|CEAM5_HUM
P40199|CEAM6_HUM
P31997|CEAM8_HUM
P13688|CEAM1_HUM
Q14002|CEAM7_HUM
Q16557|PSG3_HUMA
Q3KPI0|CEA21_HUM
P40198|CEAM3_HUM
P31809|CEAM1_MOU
Q925P2|CEAM2_MOU
P16573|CEAM1_RAT
Q63111|CEAM3_RAT
Q61400|CEAMA_MOU
Q2WEN9|CEA16_HUM
Q6UY09|CEA20_HUM
Q7Z692|CEA19_HUM
710 720 730 740 750 760 770 780 790
800 810 820 830 840 850 860 870
880 890 900 910 920 930 940 950 960
970 980 990 1000 1010 1020 1030 1040 1050
1060 1070 1080 1090 1100 1110 1120 1130 1140



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。

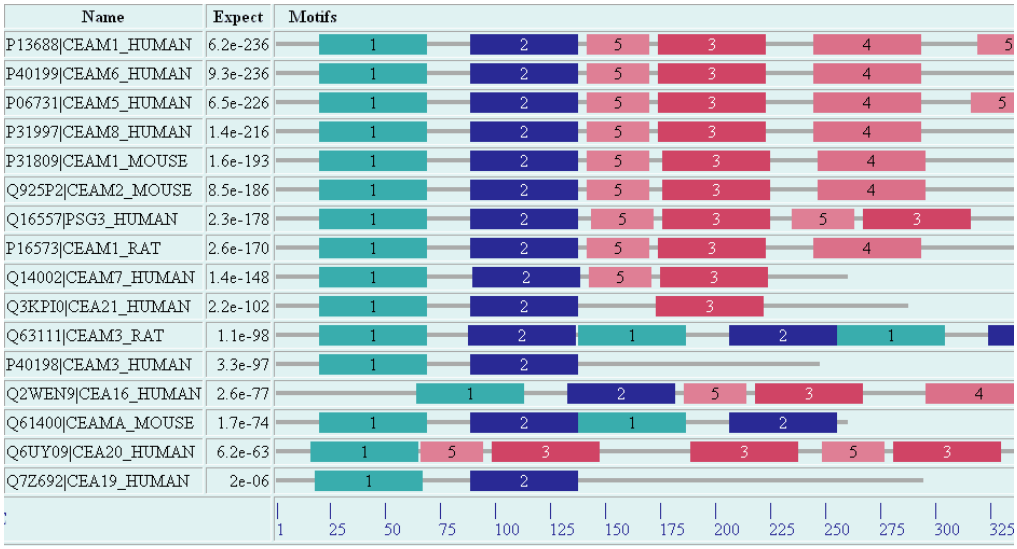
分析之二：人癌胚抗原相关蛋白质的保守基序和进化树分析

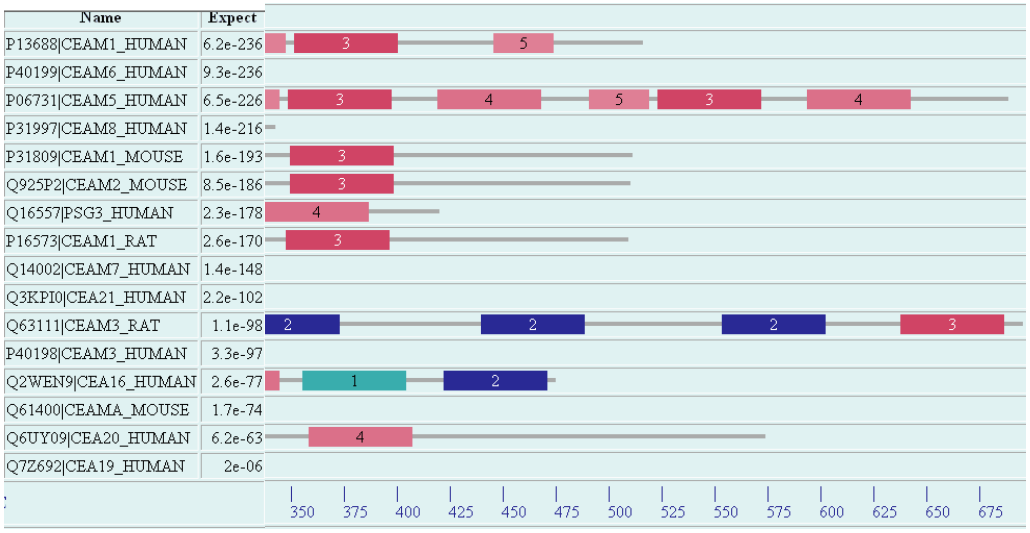
（1）MEME 分析：

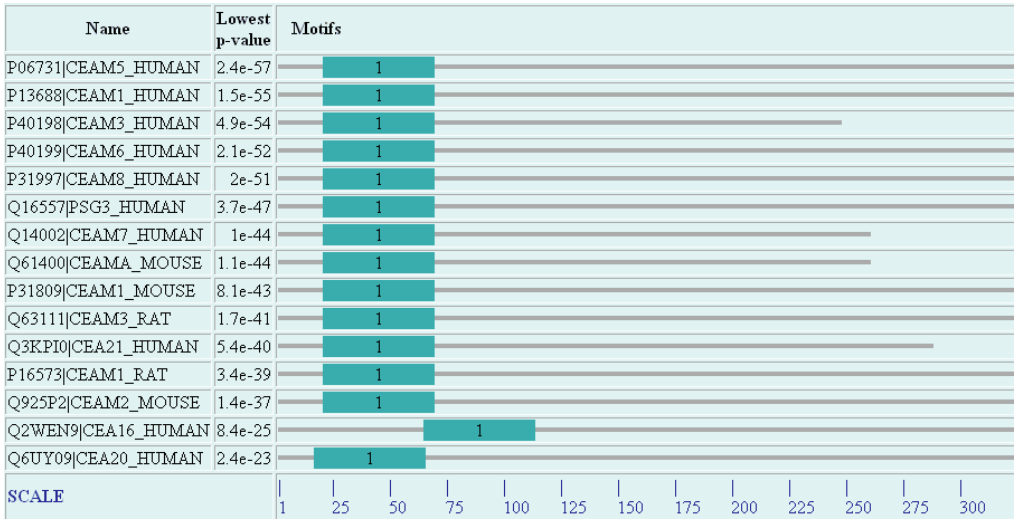
MOTIFS ./meme.html (peptide)

MOTIF	WIDTH	BEST POSSIBLE MATCH
1	50	TASLLTFWNPPTTAQVTIEAMPFNVAEGKEVLLLVHNLPQHLFGYSWYKG
2	50	PGPAYSGRETIYPNGSLLFQNVTMNDTGFYTLHMIKRDFKNEEATGQFHV
3	50	WWFNGQSLPVSDRLQLSEGNRTLTLFNVRRNDAGPYECEIWNPVSANRSD
4	50	HPGENLNLSCHAASNPPAQYFWFINGKFQQSTQELFIPNITTNNSGSYMC
5	29	PKPSITSNNSNPVEDKDAVAFTCEPETQN

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







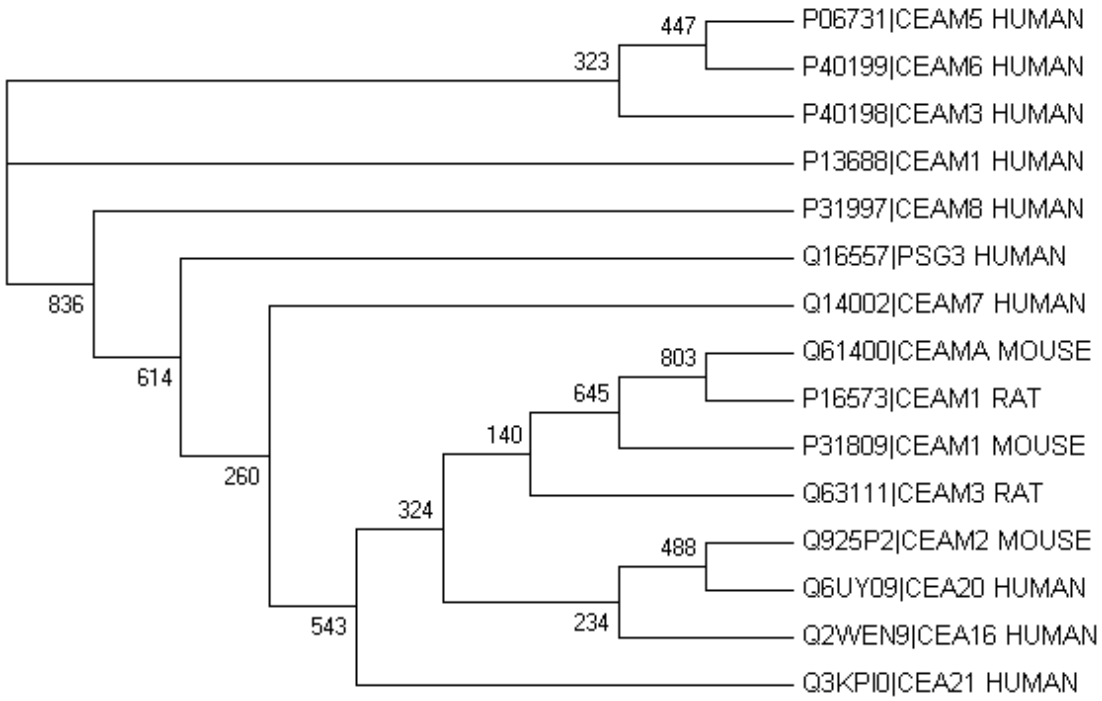
>P06731|CEAM5_HUMAN(start=21)TASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKG
>P13688|CEAM1_HUMAN(start= 21)TASLLTFWNPPTTAQLTTESMPFNVAEGKEVLLLVHNLPQQLFGYSWYKG
>P40198|CEAM3_HUMAN(start= 21)TASLLNFWNPPTTAKLTIESMPLSVAEGKEVLLLVHNLPQHLFGYSWYKG
>P40199|CEAM6_HUMAN(start= 21)TASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLAHNLPQNRIGYSWYKG
>P31997|CEAM8_HUMAN(start= 21)TASLFTFWNPPTTAQLTIEAVPSNAAEGKEVLLLVHNLPQDPRGYNWYKG
>Q16557|PSG3_HUMAN(start= 21)TALLLNFWNLPTTAQVTIEAEPTKVSKGKDVLLLVHNLPQNLAGYIWYKG

>Q14002|CEAM7_HUMAN(start= 21)TASLLTFWNLPNSAQTNIDVVPFNVAEGKEVLLVVHNESQNLYGYNWYKG
>Q61400|CEAMA_MOUSE(start= 21)TASLLTYWSPATTAQVTVEAVPPNVTADNNVLLLVHNLPQTLRVFYWYKG
>P31809|CEAM1_MOUSE(start= 21)TASLLASWSPATTAEVTIEAVPPQVAEDNNVLLLVHNLPLALGAFAWYKG
>Q63111|CEAM3_RAT(start= 21)TASLLTCWLLPTTAQVSIESLPPQVVEGENVLLHVDNLPENLIAFVWYKG
>Q3KPI0|CEA21_HUMAN(start= 21)TASLLTFWNAPTTAWLFIASAPFEVAEGENVHLSVVYLPENLYSYGWYKG
>P16573|CEAM1_RAT(start= 21)TASLLTYWSPLTTAQVTVDAVPPNVVEESSVLLLTHNLPQEFQVFYWYKV
>Q925P2|CEAM2_MOUSE(start= 21)TASLLASWSPPTTAQVTVMAFPLHAAEGNNVILVVYNMMKGVSAFSWHKG
>Q2WEN9|CEA16_HUMAN(start= 66)SWLLLSATFLNVGAEISITLEPAQPSEGDNVTLVVHGLSGELLAYSWYAG
>Q6UY09|CEA20_HUMAN(start= 17)SASLCTVWSPPAAAQLTLNANPLDATQSEDVVLPVFGTPRTPQIHGRSRE

（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：

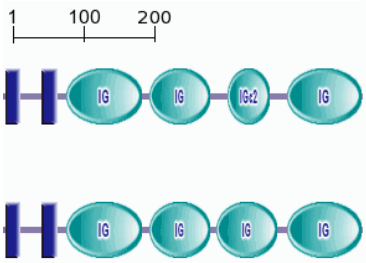


由此对基序 1 进行的进化树分析图可以初步得出上述几种蛋白之间的进化关系，结果与前面进行的序列比对结果差异较大，不仅说明比对与进化树构建两种分析途径的不同，也说明 CEA 家族成员内部之间的明显差异性，研究 CEA 家族面临的困难可能要超过以往。

分析之三：人癌胚抗原相关蛋白质的结构域和跨膜分析

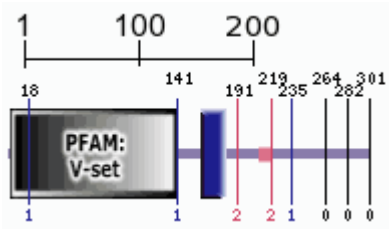
（1）SMART 分析：

>Q2WEN9|CEA16_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 16 - Homo sapiens (Human).



Name	Begin	End	E-value
transmembrane	4	23	-
transmembrane	53	75	-
IG	85	186	2.31e+00
IG	197	279	6.97e-03
IG	288	370	4.01e-03
IGc2	302	359	1.19e-05
IG	382	481	2.99e-02

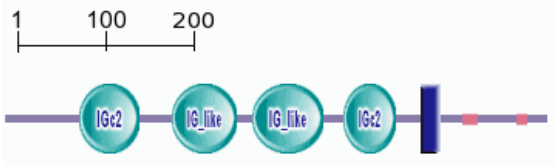
>Q7Z692|CEA19_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 19 - Homo sapiens (Human).



Name	Begin	End	E-value
Pfam:V-set	2	141	7.50e-03
transmembrane	160	182	-
low complexity	209	219	-

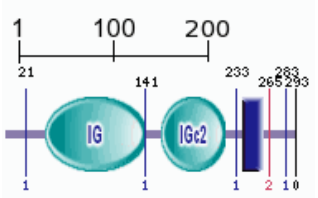
Immunoglobulin V-set domain

>Q6UY09|CEA20_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 20 - Homo sapiens (Human).



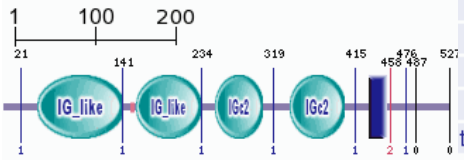
Name	Begin	End	E-value
IGc2	81	145	5.24e-07
IG_like	180	250	8.38e-06
IG_like	267	343	8.26e-05
IGc2	366	423	1.03e-05
transmembrane	449	471	-
low complexity	494	510	-
low complexity	552	564	-

>Q3KPI0|CEA21_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 21 - Homo sapiens (Human).



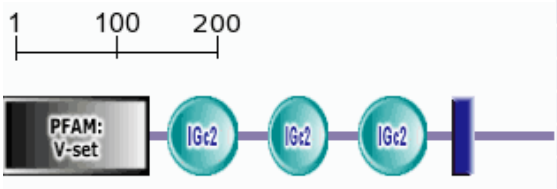
Name	Begin	End	E-value
IG	40	141	1.29e-06
IGc2	157	221	4.45e-10
transmembrane	239	261	-

>P13688|CEAM1_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 1 - Homo sapiens (Human).



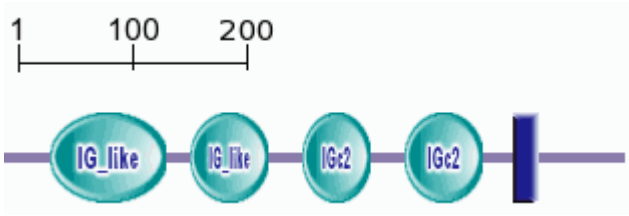
Name	Begin	End	E-value
IG_like	40	141	6.31e-01
low complexity	150	155	-
IG_like	158	234	2.96e-06
IGc2	250	306	3.06e-08
IGc2	339	403	2.68e-13
transmembrane	433	455	-

>P31809|CEAM1_MOUSE Carcinoembryonic antigen-related cell adhesion molecule 1 - Mus musculus (Mouse).



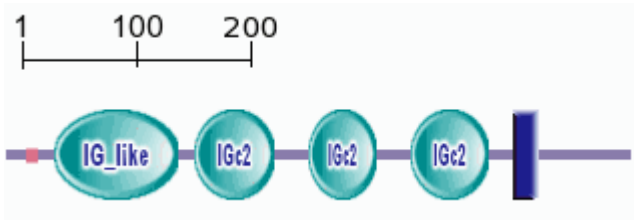
Name	Begin	End	E-value
Pfam:V-set	1	141	3.20e-15
IGc2	158	224	1.61e-07
IGc2	252	308	5.04e-09
IGc2	337	401	8.37e-15
transmembrane	426	448	-

>P16573|CEAM1_RAT Carcinoembryonic antigen-related cell adhesion molecule 1 - Rattus norvegicus (Rat)



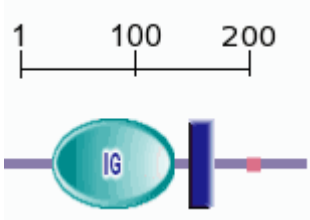
Name	Begin	End	E-value
IG_like	41	137	2.94e+00
IG_like	157	222	2.83e-03
IGc2	250	306	3.62e-10
IGc2	335	399	7.07e-13
transmembrane	426	448	-

>Q925P2|CEAM2_MOUSE Carcinoembryonic antigen-related cell adhesion molecule 2 - Mus musculus (Mouse).



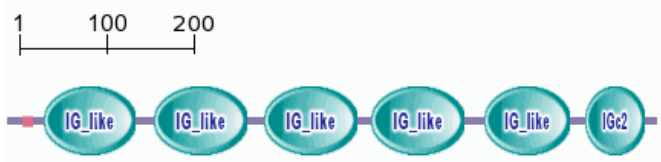
Name	Begin	End	E-value
low complexity	18	27	-
IG_like	40	143	5.41e+00
IGc2	158	224	1.99e-07
IGc2	252	308	5.04e-09
IGc2	337	401	3.28e-08
transmembrane	422	444	-

>P40198|CEAM3_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 3 - Homo sapiens (Human).



Name	Begin	End	E-value
IG	40	141	7.25e-01
transmembrane	154	176	-
low complexity	203	215	-

>Q63111|CEAM3_RAT Carcinoembryonic antigen-related cell adhesion molecule 3 - Rattus norvegicus (Rat).

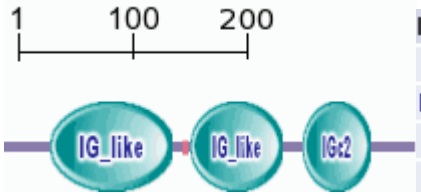


Name	Begin	End	E-value
low complexity	18	30	-
IG_like	40	139	2.15e-03
IG_like	160	261	8.33e-01
IG_like	280	381	3.94e-02
IG_like	398	499	8.46e-02
IG_like	520	613	1.39e-02
IGc2	631	695	1.30e-11

>P06731|CEAM5_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 5 - Homo sapiens (Human).

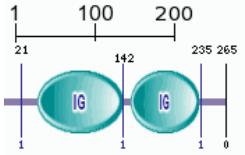
Name	Begin	End	E-value
IG_like	40	141	6.35e+00
low complexity	145	156	-
IG_like	158	234	2.37e-05
IGc2	250	306	6.65e-09
IG_like	336	412	2.64e-03
IGc2	428	484	3.56e-09
low complexity	501	512	-
IG_like	514	590	4.36e-04
low complexity	593	604	-
IGc2	606	662	9.49e-05

>P40199|CEAM6_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 6 - Homo sapiens (Human).



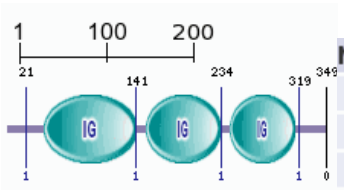
Name	Begin	End	E-value
IG_like	40	141	2.39e-01
low complexity	150	155	-
IG_like	158	234	5.93e-06
IGc2	250	306	2.13e-07

>Q14002|CEAM7_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 7 - Homo sapiens (Human).



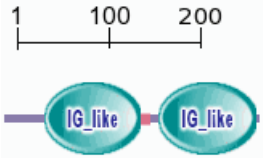
Name	Begin	End	E-value
IG	40	142	2.79e+00
IG	153	235	2.68e-04

>P31997|CEAM8_HUMAN Carcinoembryonic antigen-related cell adhesion molecule 8 - Homo sapiens (Human).



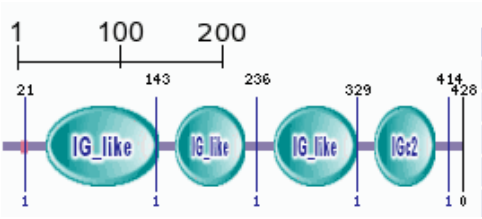
Name	Begin	End	E-value
IG	40	141	3.02e+00
IG	152	234	2.54e-05
IG	244	316	1.48e-06

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



Name	Begin	End	E-value
IG_like	42	141	1.23e+01
low complexity	142	151	-
IG_like	160	261	3.97e+01

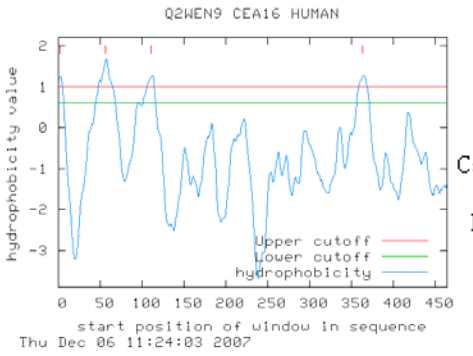
>Q16557|PSG3_HUMAN Pregnancy-specific beta-1-glycoprotein 3 - Homo sapiens (Human).



Name	Begin	End	E-value
low complexity	18	25	-
IG_like	40	145	1.79e+00
IG_like	160	224	9.97e-02
IG_like	253	329	1.23e-03
IGc2	345	401	2.32e-08

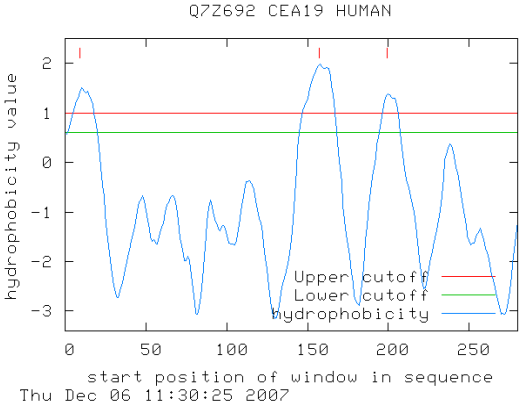
SMART 分析可以初步解释在前面序列比对中出现的重复点阵区应该是免疫球蛋白结构域区，也可以一定程度上解释出 CEA5 与 CEA3 的综合相似程度高，还可以看出 CEA 家族成员普遍具有免疫球蛋白结构域，从这一点出发，可以利用已知的免疫球蛋白三维结构来推测未知结构的 CEA 家族蛋白。

（2）TopPred 分析：



Candidate membrane-spanning segments:

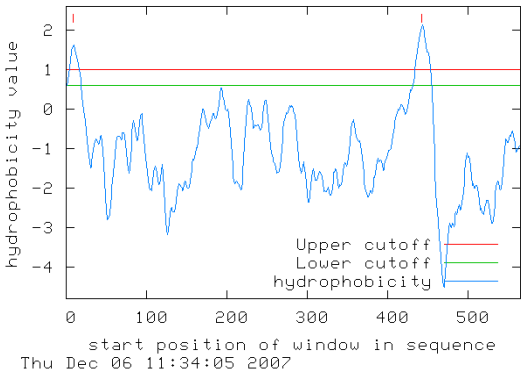
Helix	Begin	-	End	Score	Certainity
1	3	-	23	1.275	Certain
2	57	-	77	1.692	Certain
3	112	-	132	1.285	Certain
4	364	-	384	1.293	Certain



Candidate membrane-spanning segments:

Helix	Begin	-	End	Score	Certainity
1	10	-	30	1.499	Certain
2	158	-	178	1.975	Certain
3	200	-	220	1.390	Certain

Q6UY09 CEA20 HUMAN

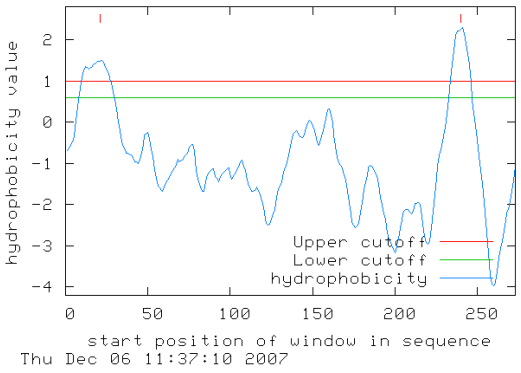


Thu Dec 06 11:34:05 2007

Candidate membrane-spanning segments:

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

Q3KPI0 CEA21 HUMAN

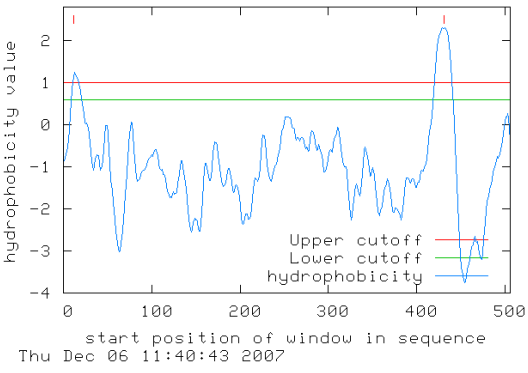


Thu Dec 06 11:37:10 2007

Candidate membrane-spanning segments:

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

P13688 CEAM1 HUMAN

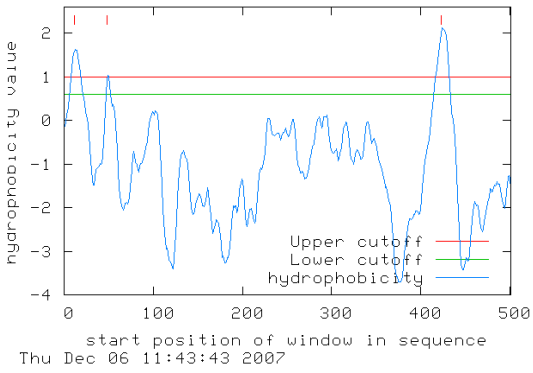


Thu Dec 06 11:40:43 2007

Candidate membrane-spanning segments:

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

P31809 CEAM1 MOUSE

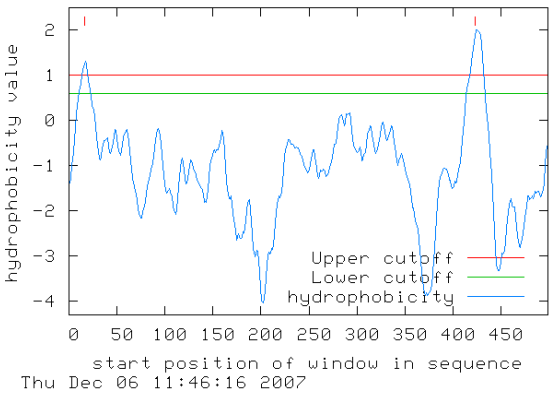


Thu Dec 06 11:43:43 2007

Candidate membrane-spanning segments:

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

P16573 CEAM1 RAT

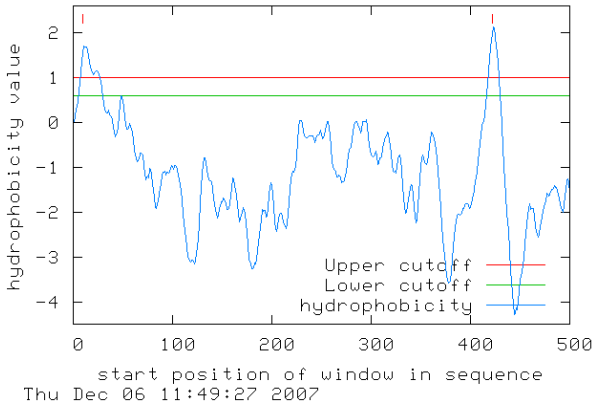


Thu Dec 06 11:46:16 2007

Candidate membrane-spanning segments:

Helix	Begin		End	Score	Certainity
1	17	-	37	1.316	Certain
2	424	-	444	2.025	Certain

Q62SP2 CEAM2 MOUSE

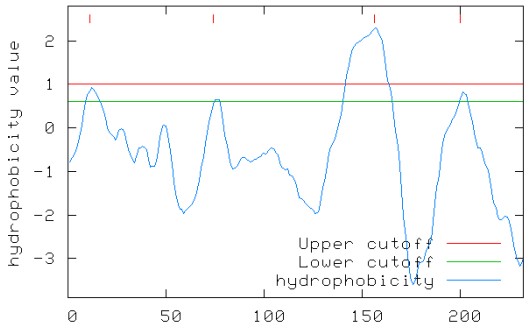


Thu Dec 06 11:49:27 2007

Candidate membrane-spanning segments:

Helix	Begin		End	Score	Certainity
1	11	-	31	1.701	Certain
2	423	-	443	2.138	Certain

P40198 CEAM3 HUMAN

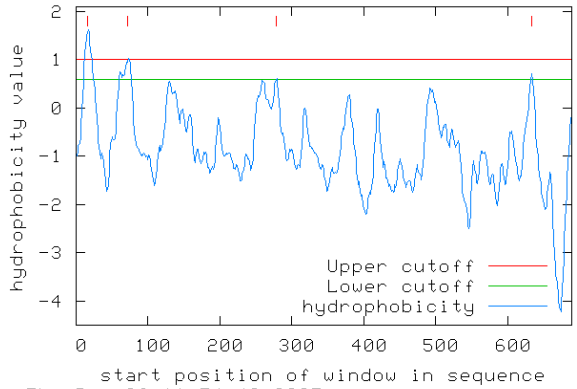


Thu Dec 06 11:52:04 2007

Candidate membrane-spanning segments:

Helix	Begin		End	Score	Certainity
1	12	-	32	0.927	Putative
2	75	-	95	0.652	Putative
3	157	-	177	2.297	Certain
4	201	-	221	0.828	Putative

Q63111 CEAM3 RAT

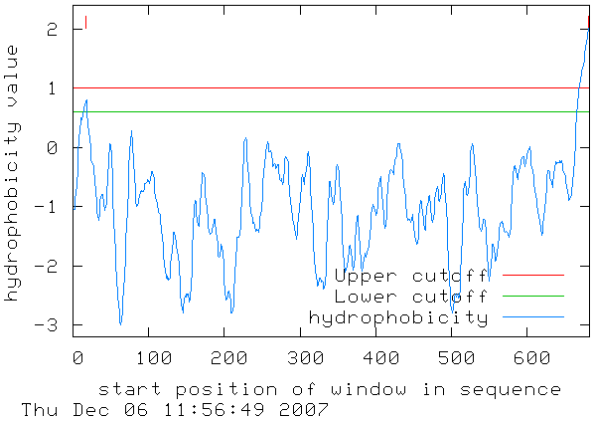


Thu Dec 06 11:54:43 2007

Candidate membrane-spanning segments:

Helix	Begin		End	Score	Certainity
1	17	-	37	1.616	Certain
2	73	-	93	1.023	Certain
3	279	-	299	0.623	Putative
4	634	-	654	0.716	Putative

P06731 CEAM5 HUMAN

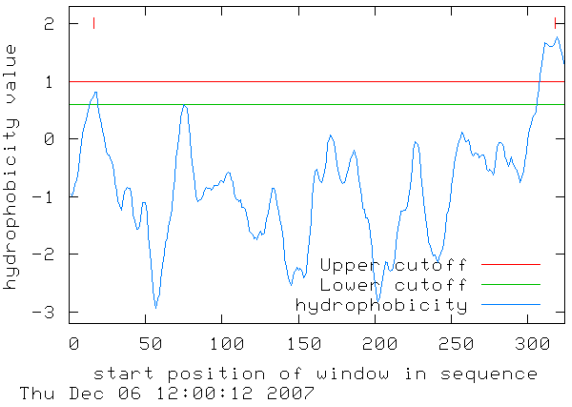


Thu Dec 06 11:56:49 2007

Candidate membrane-spanning segments:

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

P40199 CEAM6 HUMAN

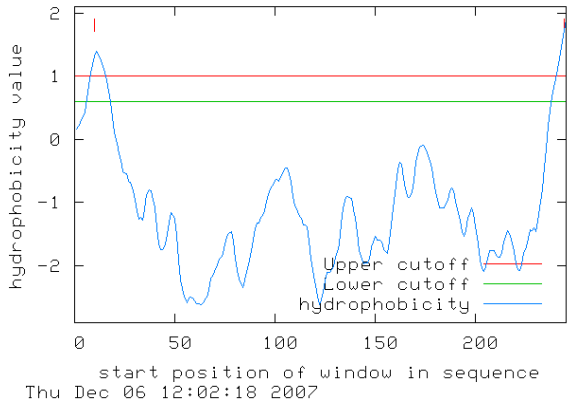


Thu Dec 06 12:00:12 2007

Candidate membrane-spanning segments:

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

Q14002 CEAM7 HUMAN

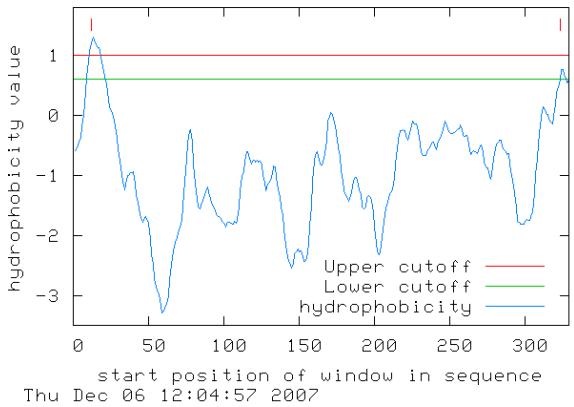


Thu Dec 06 12:02:18 2007

Candidate membrane-spanning segments:

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

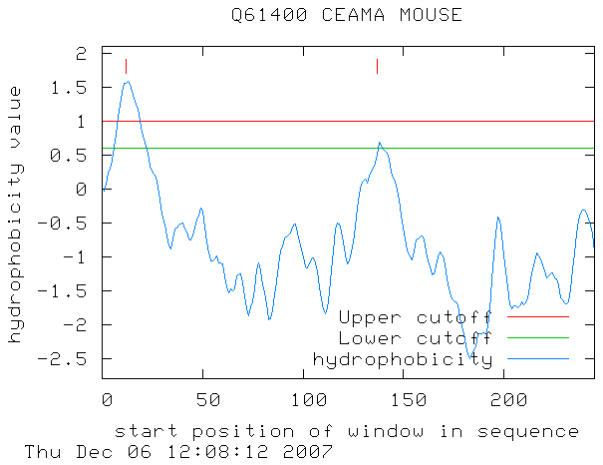
P31997 CEAM8 HUMAN



Thu Dec 06 12:04:57 2007

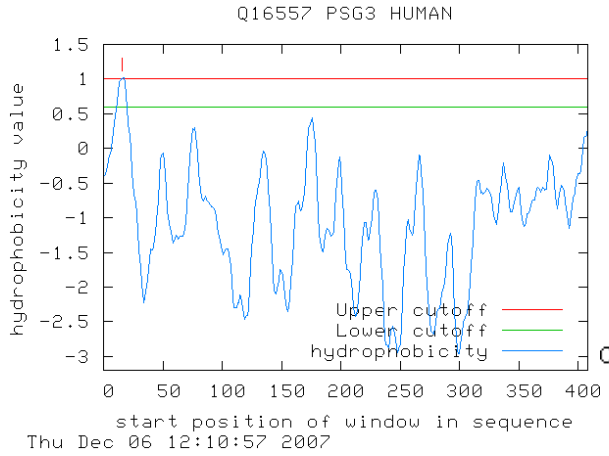
Candidate membrane-spanning segments:

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



Candidate membrane-spanning segments:

Helix	Begin - End	Score	Certainity
1	13 - 33	1.591	Certain
2	138 - 158	0.686	Putative



Candidate membrane-spanning segments:

Helix	Begin - End	Score	Certainity
1	17 - 37	1.016	Certain

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。