



Structure Prediction of AGGF1

Xu Ke

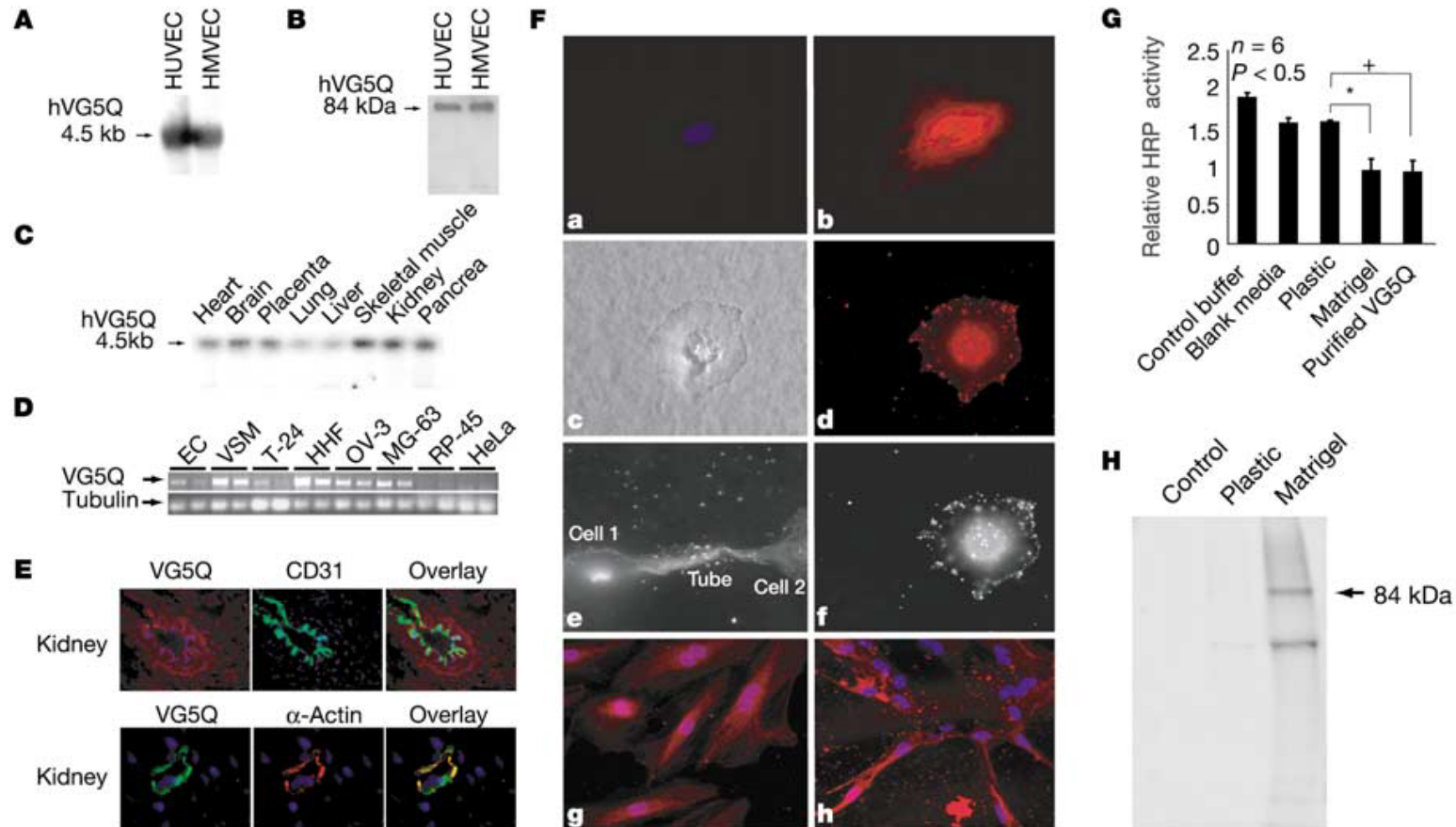
Hu Xiaomin

Wu Qihui

Yang Xi

Background: The Biology of AGGF1

AGGF1: angiogenic factor with G patch and FHA domains 1



The Structure of AGGF1



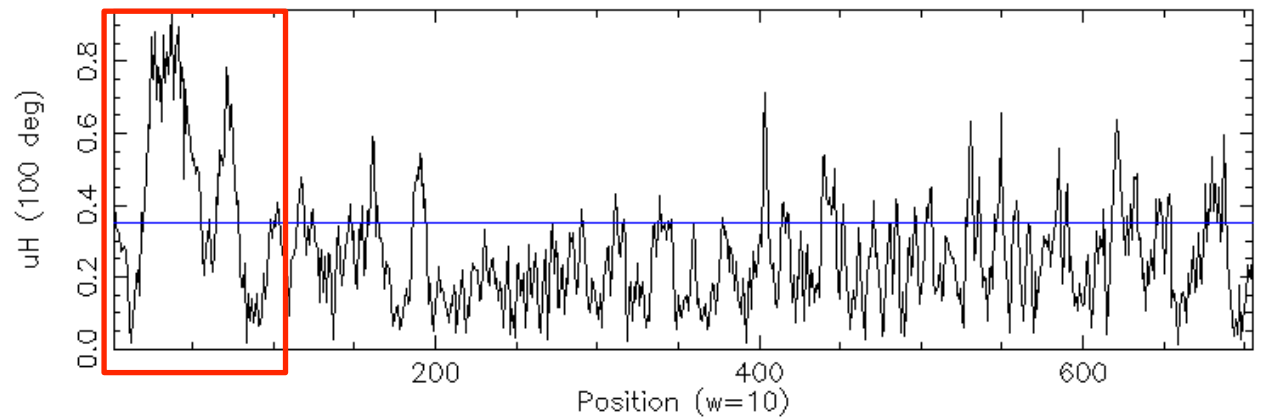
- Sequence Analysis
- Motif Prediction
- Secondary Structure Analysis
- Structure Prediction

Sequence Analysis

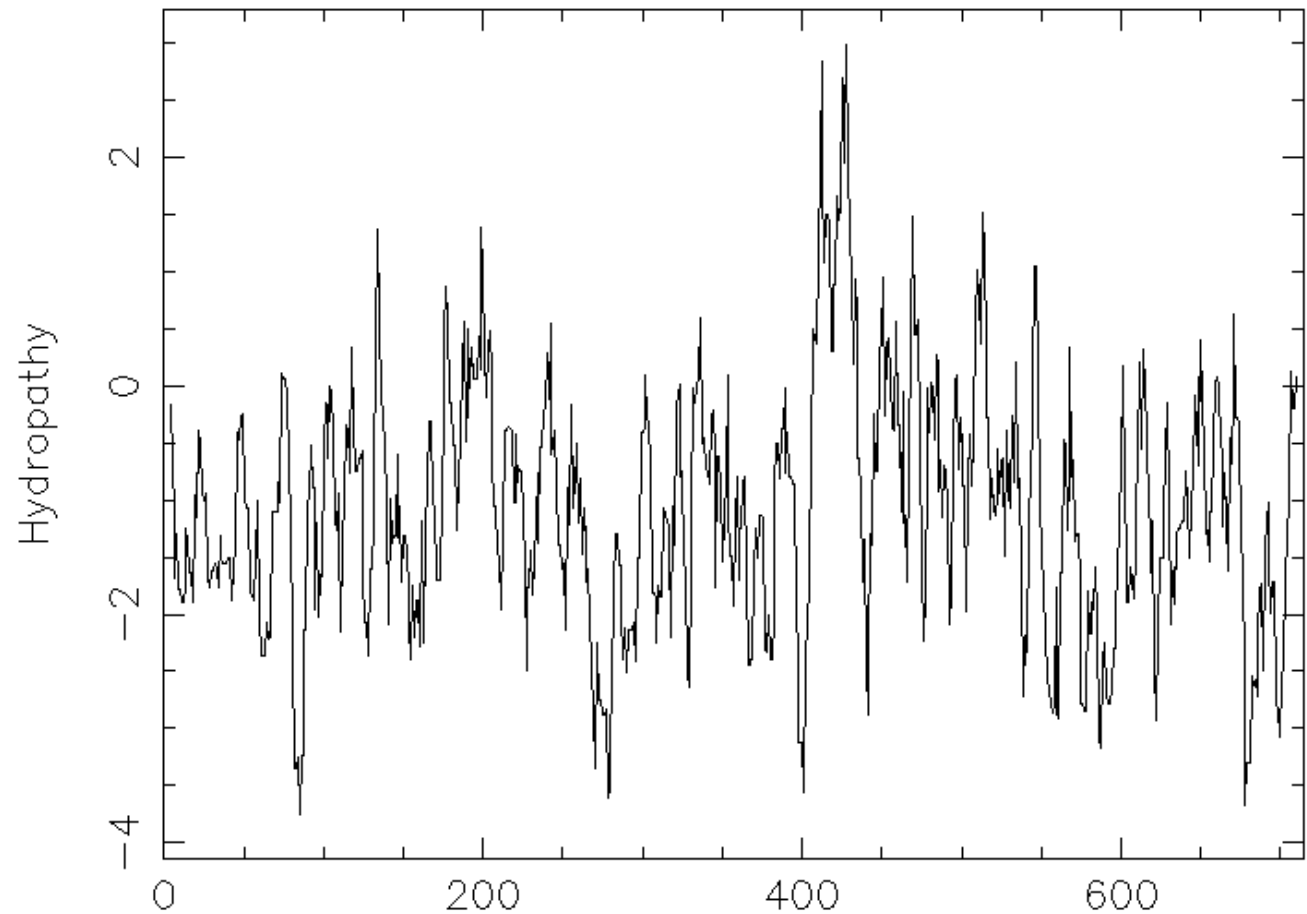
Molecular weight = 80977.11 Residues = 714
Average Residue Weight = 113.413 Charge = -19.5
Isoelectric Point = 5.1109
A280 Molar Extinction Coefficient = 60450
A280 Extinction Coefficient 1mg/ml = 0.75
Probability of expression in inclusion bodies = 0.497

Property	Residues	Number	Mole%
Tiny (A+C+G+S+T)	189	26.471	
Small (A+B+C+D+G+N+P+S+T+V)	350	49.020	
Aliphatic (A+I+L+V)	145	20.308	
Aromatic (F+H+W+Y)	69	9.664	
Non-polar (A+C+F+G+I+L+M+P+V+W+Y)	289	40.476	
Polar (D+E+H+K+N+Q+R+S+T+Z)	425	59.524	
Charged (B+D+E+H+K+R+Z)	247	34.594	
Basic (H+K+R)	119	16.667	
Acidic (B+D+E+Z)	128	17.927	

Hydrophobic Analysis



Hydropathy Analysis




Signal Peptide Prediction

No signal Peptide is found in several programs.
It **may** contain a NLS.

```
MASEAPSPRSPPPPTSPEPELAQLRRKVEKLERELRSCKRQVREIEKLLHHTERLYQNA  
ESNNQELRTQVEELSKILQRGRNEDNKKSDVEVQTENHAPWSISDYFYQTYINDVSLPNK  
VTELSDQQDQAIETSILNSKDHLQVENDAYPGTDRTENVKYRQVDHFAASNSQEPASALAT  
EDTSLEGSSLAESLRAAAEAASVSQTGFSDYDENTGLYFDHSTGFYDSENQLYDPSTGIY  
YYCDVESGRYQFHRSRVDLQPYPTSSTKQSKDKKLKKRKDPDSSATNEEKDLNSEDQKAF  
SVEHTSCNEEENFANMKKKAKIGIHHKNSPPKVTVPVPTSGNTIESPLHENISNSTSFKDEK  
IMETDSEPEEGEITDSQTEDSYDEAITSEGNVTAEDSEDEDEDKIWPPCIRVIVIRSPVL  
QIGSLFIITAVNPFATIGREKDMHTLRIPVGVSKFHAEIYFDHDLQSYVLVDQGSQNGT  
IVNGKQILQPKTKCDPYVLEHGDEVKIGETVLSFHIHPGSDTCDGCEPGQVRAHLRLDKK  
DESFVGPVTLKKEEKELERRKELKIRVKYGLQNTYEDEKTLKNPKYKDRAGKRREQVGS  
EGTFQRDDAPASVHSEITDSNKGRKMLEKMGWKKGEGLGKGGGMKTPIQQLRRTHAGL  
GTGKPFSSFEDVHLLQNKKNKNWDKARERFTENFPETKPKQKDDPGTMPWVKGTLE
```

Signal Peptide Prediction



Predicted Localization: Not Secreted, Not Nuclear, Cytoplasmic.

But it is enriched in nuclear, and also can be secreted.

Explanation:

Noncanonical secreted pathway exist in many growth factors, such as FGF family.

AGGF1 can bind to other factors when it is transported into nuclear.

Phosphorylation Site Prediction

Description	Position	Score	Sequence	Profile
a	269	0.55	QFHSRVLDLQPYPTSSTKQSKDKKLKKRKRKDPDSSATNEEKDL	0.72
	329	0.67	ENFANMKKKAKIGIHHKNSPPKVTVPVPTSGNTIESPLHENISN	1.10
	334	0.24	MKKKAKIGIHHKNSPPKVTVPVPTSGNTIESPLHENISNSTSFK	0.38
	656	0.53	LGKDGGGMKTPIQLQLRRTHAGLGTGKPFSSFEDVHLLQNKNK	0.60
	666	0.19	PIQLQLRRTHAGLGTGKPFSSFEDVHLLQNKNKKNWDKARERF	0.67
	712	0.18	PETKPQKDDPGTMPWVKGTLE*	0.40

Summary

Submitted sequences: 1
Submitted S/T residues: 109
Number of predicted sites: 6

Phosphorylation Site Prediction

<u>SS</u>	<u>MS</u>	<u>human</u>	<u>mouse</u>
0	3	S7-p _MASEAPsPPRsPPP	S7 _MASEAPSPSPSPSP
0	3	S11-p EAPsPPRsPPPPTsP	S12 APSPSPSPSPPPASP
0	2	S17-p RsPPPPTsPEPELAQ	S18 PSPPPPASPEPELAQ
0	1	D148-ca DHLQVENdAYPGTDR	- under review
0	1	- under review	S308-p SVEYINCSEDEHSGN
0	3	S543-p RLDRKDEsFVGPTLs	P540 RLDRKDEPLVGPALS
0	3	S550-p sFVGPTLsKEEKELE	S547 PLVGPALSKEEKELE
0	1	K551 FVGPTLsKEEKELER	K548-a LVGPALSKEEKELER
0	1	S615-p DAPASVHsEITDSNK	S612 DAPASVHSEITDSNK
0	1	K664-a HAGLGTGkPSSFEDV	K661 HAGLGTGKLSSIDDV
0	1	K679-a HLLQNKNkKNWDKAR	K676 HLIQNKSkkHWDKAR

Secondary Structure Analysis

20种氨基酸的Chou-Fasman参数

α螺旋参数			β折叠参数		
氨基酸残基	P _α	能力强弱	氨基酸残基	P _β	能力强弱
Glu(-)	1.51	H	Val	1.70	H
Met	1.45	H	Ile	1.60	H
Ala	1.42	H	Tyr	1.47	H
Leu	1.21	H	Phe	1.38	H
Lys(+)	1.16	H	Trp	1.37	h
Phe	1.13	H	Leu	1.30	h
Gln	1.11	H	Cys	1.19	h
Trp	1.08	H	Thr	1.19	h
Ile	1.08	H	Gln	1.10	h
Val	1.06	H	Met	1.05	h
Asp(-)	1.01	I	Arg(+)	0.93	i
His(+)	1.00	I	Asn	0.89	i
Arg(+)	0.98	i	His(+)	0.87	i
Thr	0.83	i	Ala	0.83	i
Ser	0.77	i	Ser	0.75	b
Cys	0.70	b	Gly	0.75	b
Tyr	0.69	b	Lys(+)	0.74	b
Asn	0.67	b	Pro	0.55	b
Pro	0.57	B	Asp(-)	0.54	b
Gly	0.57	B	Glu(-)	0.37	B

Secondary Structure Analysis

AGGF1 is a helix-rich protein

```
      . 10 . 20 . 30 . 40 . 50
maseapsprspppptspepelaqlrrkveklerelrschrqvreiekll
helix HH                HHHHHHHHHHHHHHHHHHHHH HHHHHHHHHHH
sheet
turns      T  TT                      T
coil      CCC CC  CCCCCCCCC

      . 60 . 70 . 80 . 90 . 100
hhterlyqnaesnnqelrtqveelskilqrgrnednkksdvevqtenhap
helix HHHHHH          HHHHHHHHHHHHHHHHHHHHH HHHHHHHH
sheet
turns      T                      T TTTT          T  T
coil      CCCCCCCC          CC C  CC          CC
      . 110 . 120 . 130 . 140 . 150
wsisdyyfyqyyndvslpnkvtelsdqddqaietsilnksdhlqvenday
helix
sheet      EEE          EE          EEE          EEEE
turns T TTTTT  TTTTT          T  T          TT
coil C          CCCCCC          C C CC          CC
      . 160 . 170 . 180 . 190 . 200
pgtdrtenvkyrqvdhfasnsqepasalatdtslegsslaeslraaaaa
helix
sheet      E          EEEEE
turns T  TTTTT
coil C CCC          CCCCCCCC          C
      . 210 . 220 . 230 . 240 . 250
avsqtgfsydgntglyfdhstgfydsenqlyydpstgiyyycdvesgry
helix HHHH
sheet      EEE          EEEEEEE
turns      TTTTT  TTTTT  TTTTTTTT TTTT          TTTTT
coil      CCCC  C          CC          C
```

```
      . 260 . 270 . 280 . 290 . 300
qfhsrvdlqpyptsstksqskdkklkkdkdkdpdssatneekdlnsedqkaf
helix
sheet      E          E
turns TT TTTT          T  T          TT  T
coil      C CCC CC          CC CCC
      . 310 . 320 . 330 . 340 . 350
svehtscneeeenfamkkkaki gi h h k n s p p k v t v p t s g n t i e s p l h e n i
helix HHH  HHHHHHHHHHHHHHHHHHHHH
sheet
turns
coil      CCCC          C  CC          CCCCC CCCCC
      . 360 . 370 . 380 . 390 . 400
snstsfkdekimetdsepegeitdsqtedsydeaitsegnvtaedsede
helix
sheet
turns TT          T
coil      CCC          CCC          CCCC C          CCCCC
      . 410 . 420 . 430 . 440 . 450
dedkiwppcirvivirspvlqigslfiitavnpatigrekdmehltripe
helix HHH
sheet      EEEEEEEEEEEEEEEEEEEEEEE
turns TT  TT
coil      CC          CCCCC
      . 460 . 470 . 480 . 490 . 500
vgvskfhaei y f d h d l q s y v l v d q g s q n g t i v n g k q i l q p k t c d p y v l e
helix HHHHHHHHHHHHHHHHHHHHH
sheet
turns      T  TT          TT          TT          EEEEE  EEEEE  EEEEE
coil      C          CCCC
```

Secondary Structure Analysis

```

      . 510      . 520      . 530      . 540      . 550
      hgdevkigetvlsfhihpgsdtcdgcepqvrahlrldkkdesfvqptls
helix  HHHHHHHHHHHHHH          HHHHHHHHH          H
sheet          EEE                      EEEE
turns          TTTTTTTTTT TT          TTTT
coil          CC          C          CCCC

      . 560      . 570      . 580      . 590      . 600
      keekelerrkelkkirvkyqlnteyedektlknkpkdragkrreqvgs
helix  HHHHHHHHHHHHHHHHH          HHHHHHHHH          HHHH
sheet          EEE                      EEE
turns          T                      TT          T
coil          CCCCC          C          CCC          CCC

      . 610      . 620      . 630      . 640      . 650
      egtfqrddapasvvhseitdsnkgrkmlkmgwkkgeglgkdgggmktpiq
helix          HH HH          H HHHHHHHHHHHHH
sheet  EE          EE                      EEEE
turns  T TT          T T          TTTT
coil  CC          CCCCC          C C          CCCCC          CCC

      . 660      . 670      . 680      . 690      . 700
      lqlrrthaglgtgkpsfedvhlqknknkndkarerftenfpetkpkqk
helix          HHHHHHHHH          HHHHH
sheet  EEEEE          E
turns  T T TT          TTTTT          TT          TTT TTT
coil  CC C CCCCC          C          CCCC          C

      . 710
      ddpgtmpwvkgtle
helix          H
sheet  E E
turns  T TT          T
coil  C CC          CCCC

```

Residue totals: H:286 E:104 T:
 146 C:178
 percent: H: 41.0 E: 14.9 T: 20.9
C: 25.5

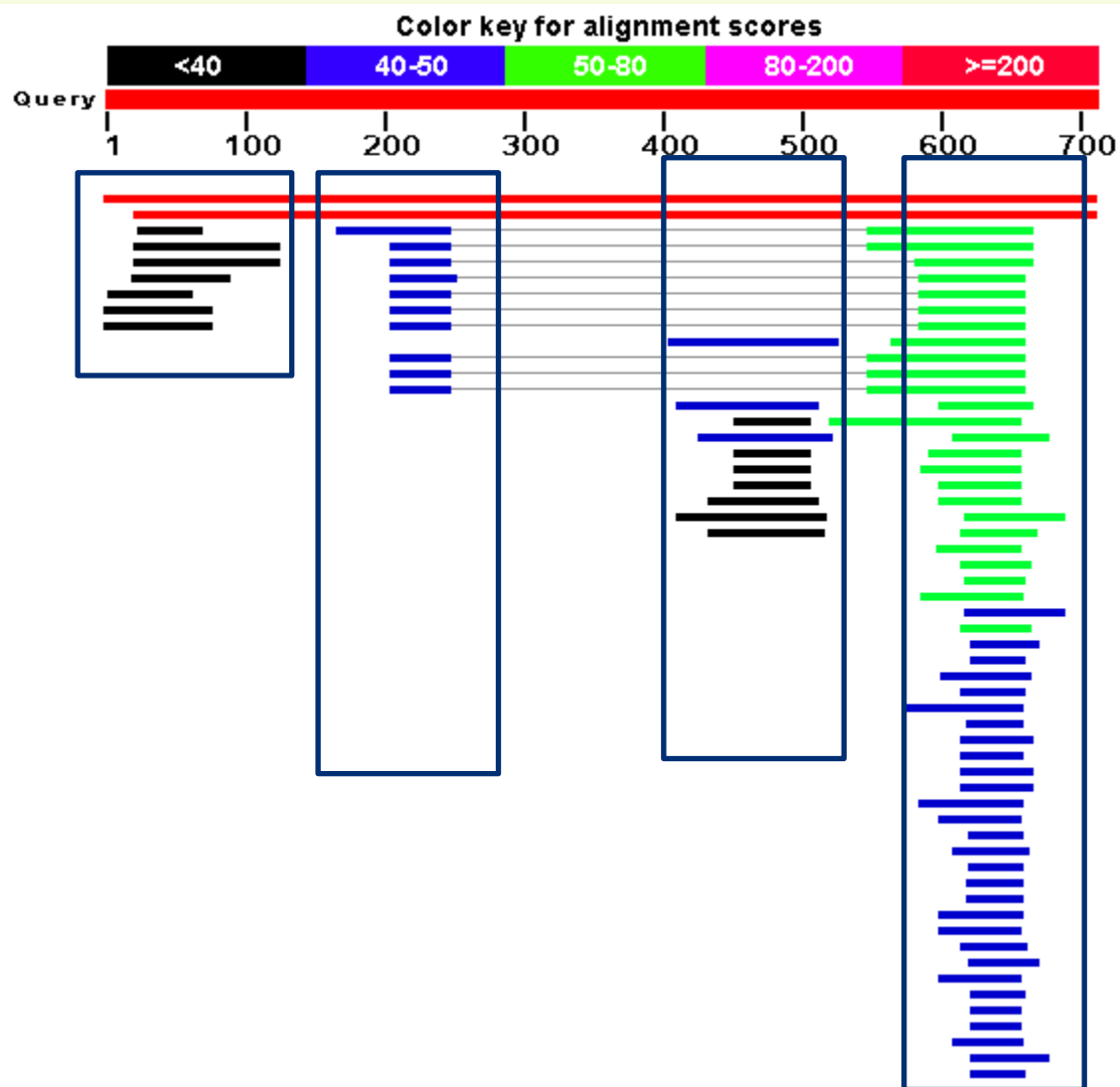
It is still very difficult to predict
 the loop structure.

Structure Prediction

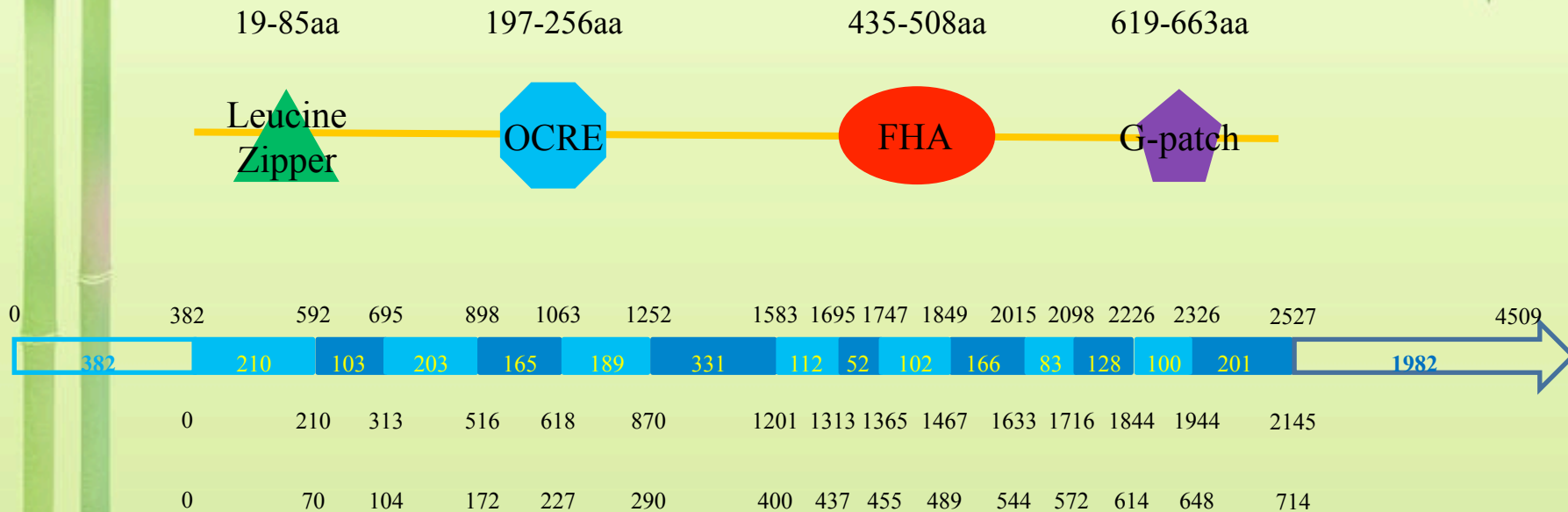


- ◆ The Blast Result
- ◆ Domain Analysis
- ◆ The Brief Analysis of Coiled-coil Domain
- ◆ The Comparative Modeling of FHA Domain

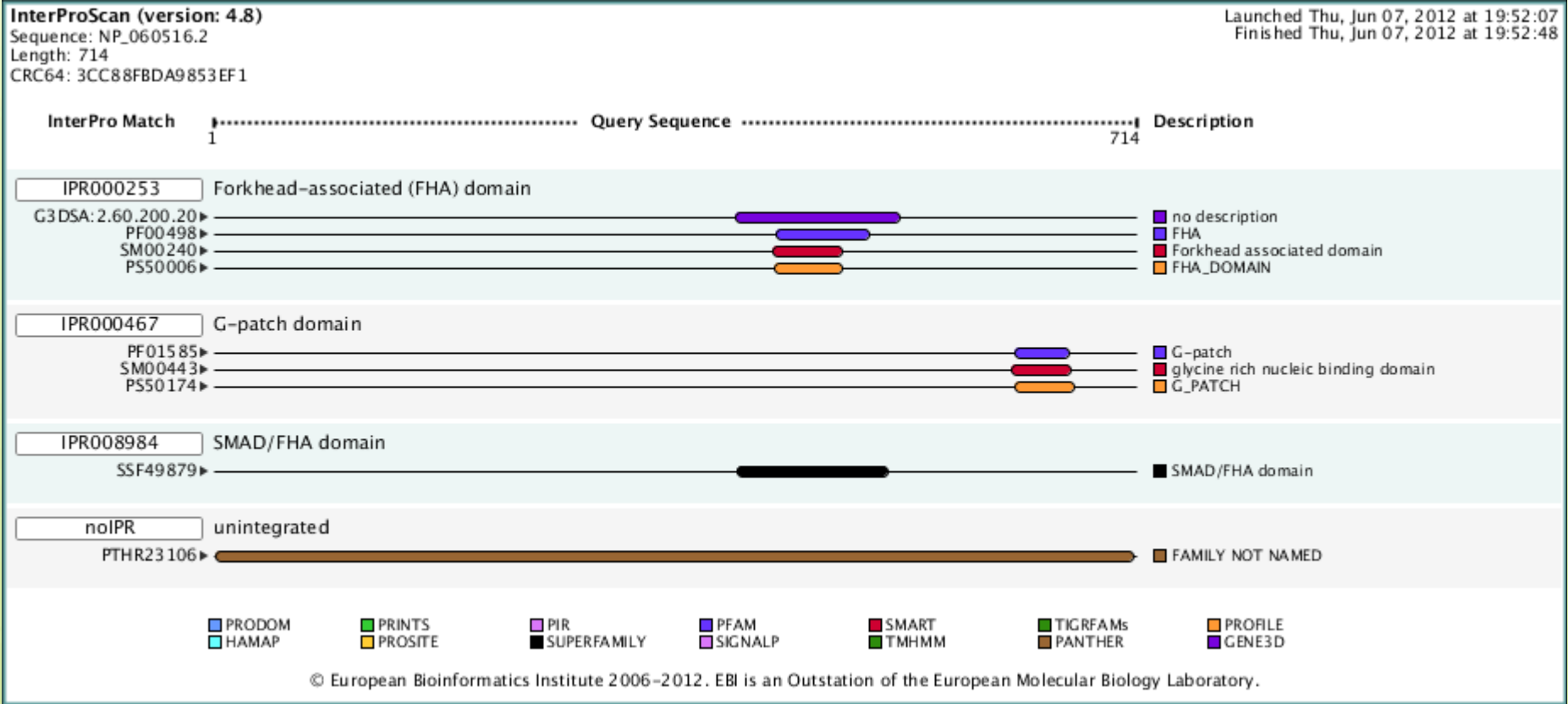
The Blast Result



The Unique Structure of AGGF1



Domain Analysis



Domain Analysis

Prediction 9; Score = 2.885; Domains = 5; Boundaries = 49 111 146 262 339 401 512 558 598



Prediction 0; Score = 2.735; Domains = 5; Boundaries = 57 146 262 374 512 540 621

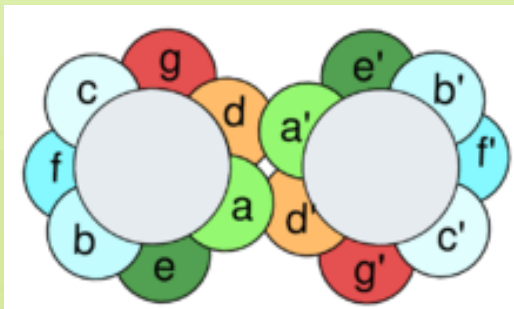
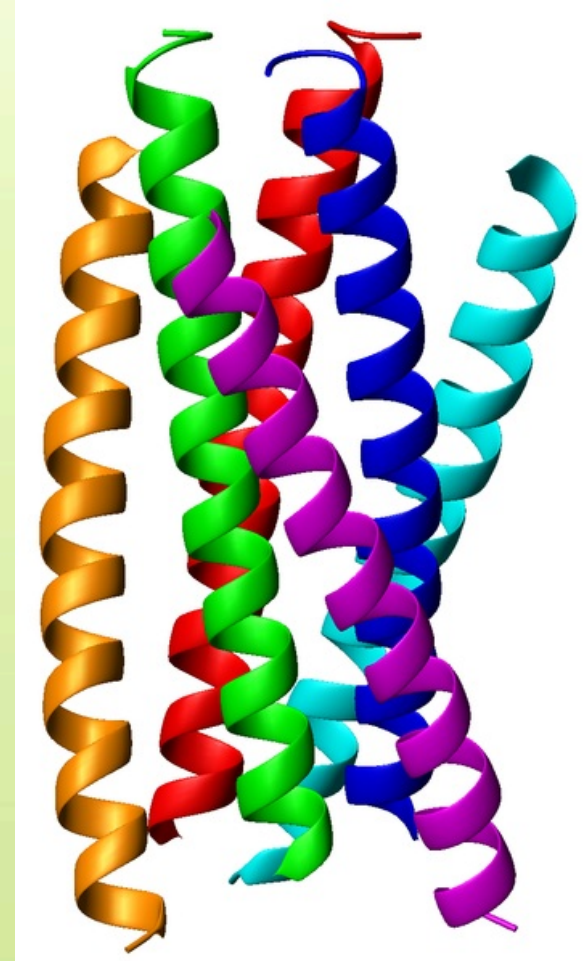
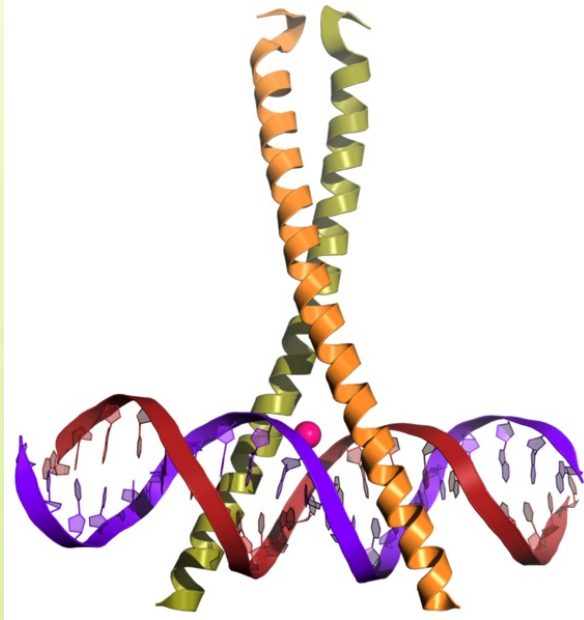


Scooby-domain prediction

To Select a Model

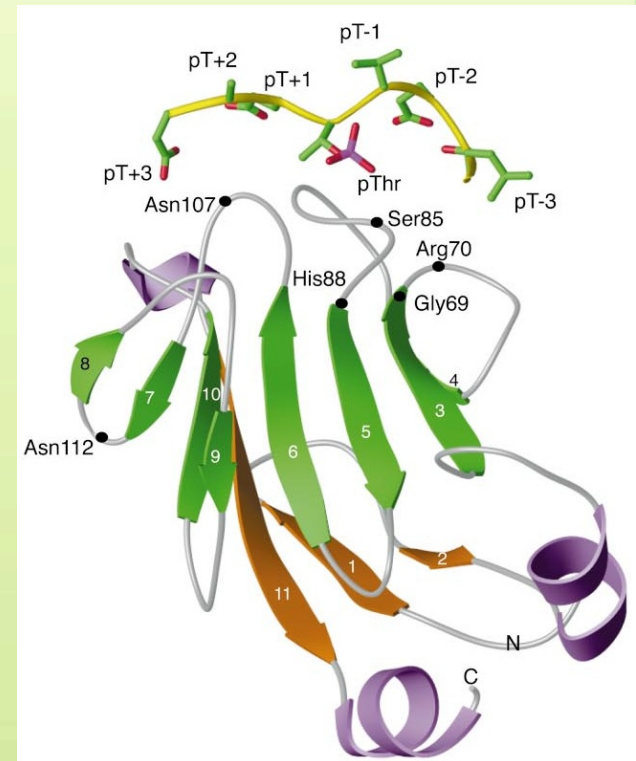
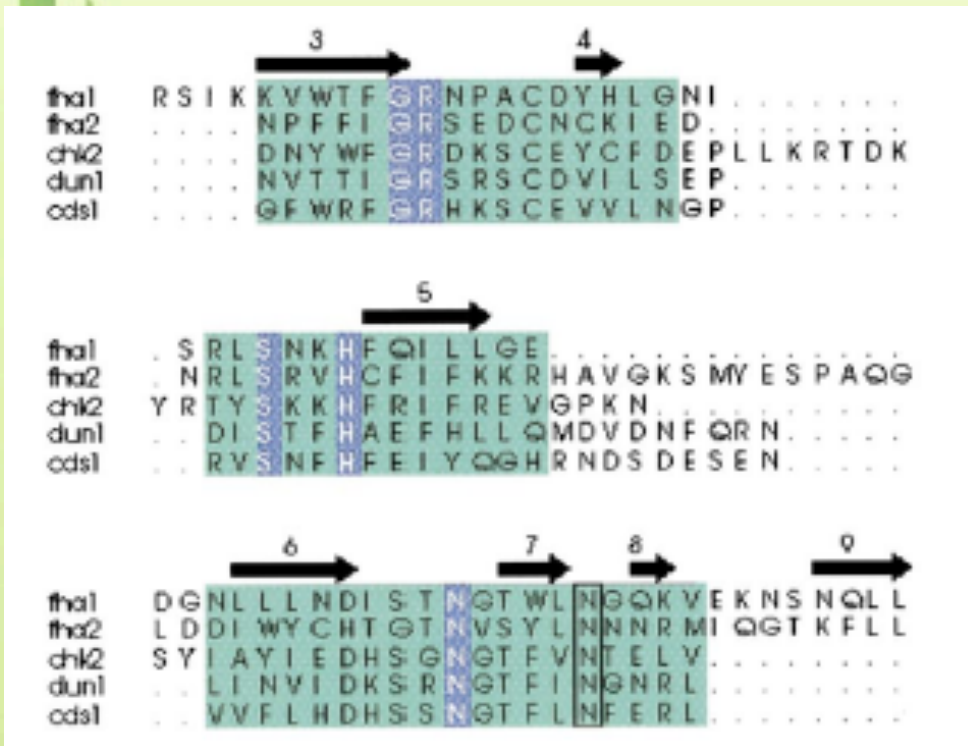
Model	Rel.	Provider	Type	Templates	%Seq id	from	to
[Show]		MODBASE	SC	1mzkA ↗	35%	433	510
[Show]		MODBASE	SC	2kkIA ↗	32%	387	515
[Show]		SWISSMODEL	SC	2lc1A ↗	30%	432	517
[Show]		SWISSMODEL	SC	2kb3A ↗	29%	410	518
[Show]		MODBASE	SC	1mzkA ↗	28%	407	516
[Show]		MODBASE	SC	1mzkA ↗	27%	410	522
[Show]		MODBASE	SC	2kfuA ↗	26%	357	529
[Show]		MODBASE	SC	1g6gA ↗	26%	407	516
[Show]		MODBASE	SC	2ff4A ↗	25%	387	520
[Show]		MODBASE	SC	1uhtA ↗	25%	406	525
[Show]		MODBASE	SC	1g3gA ↗	24%	407	520
[Show]		MODBASE	SC	1lgpA ↗	21%	408	516
[Show]		MODBASE	SC	1gxcA ↗	21%	416	521
[Show]		MODBASE	SC	1lgpA ↗	19%	409	527
[Show]		MODBASE	SC	2jeeA ↗	18%	30	86
[Show]		MODBASE	SC	1yjmA ↗	17%	432	517
[Show]		MODBASE	SC	1dmzA ↗	17%	407	554
[Show]		MODBASE	SC	3euhA ↗	15%	19	231
[Show]		MODBASE	SC	1wlnA ↗	13%	412	523
[Show]		MODBASE	SC	2ap3A ↗	13%	10	159
[Show]		MODBASE	SC	2pftA ↗	10%	19	384

The Brief Analysis of Coiled-coil Domain



From Wiki

The Comparative Modeling of FHA Domain



The Comparative Modeling of FHA Domain



2lc1: *Mycobacterium tuberculosis* Rv0020c

The Comparative Modeling of FHA Domain



SWISS Model

Introduction to G-patch Domain

Secondary Struct.	... HHHHHHHHHHHH ... HHHHHHHH ...	
Luca15_Hs_1244404	743 HSNIGNKMLQAMGWREGSGLGRKCQ-GITAPIEAQVR-----LKGAGLGAKGS	789 \
S1-1_Rr_1514971	780 SDNIGSRMLQAMGWKEGSGLGRKKQ-GIVTPIEAQTR-----VRGSGLGARGS	826
DXSB237E_Hs_1184064	317 SDNIGSRMLQAMGWKEGSGLGRKKQ-GIVTPIEAQTR-----VRGSGLGARGS	363 RRM+C4+RRM+ZNF+GP
T08B2.5_Ce_1947021	598 ESNIGNRLLQKSMGWKEGQGLVGHKAQ-GIVNPIEAERF-----VQGAGLGAAGS	644
g16_Hs_3559787	1051 TSSKGGCVQQTGWRRKGTGLVGYGHP-GLASSEEAEGR-----MRGFSVGSASGR	1097
Y55D9A.2b_Ce_3881118	474 SGNVGFKLLKSMGWSEGGQLGKKEQ-GHVEPVATEVK-----NNRKCLGANEK	520
AC57A7.13_Sp_2104448	490 TSNVSMKMLNSMGWNKSGSLGTNEN-GIKEAIQPTMY-----LPGVGLGNKGS	536 /
F4F15.230_At_4678941	1015 ADNVGHKLLSMLGWKEGEGIGSSRK-GMADPIMAGDV-----KTNNLGVGASAP	1074 \
F23858_1_Hs_2988397	525 VENIGYQMLMKMGWKEGEGGLGSEGQ-GIKNPVNKGTGTT-----TVDGAGFGIDRP	572 2*SWAP+GP
KIAA0365_Hs_2224671	868 DKNLGFQMLQKMGWKEGHGLGSLGK-GIRSRSACTQQ-----AAWGGSGWGLS	914
EPROT_Hs_2058691	809 EENKGHQMLVKMGWSGSGGLGAKEQ-GIQDPIKGGDV--RDKWDQYKGVGVALD	859 \
F58G11.5_Ce_3877941	646 YDNKGAQLMAKMGWG-GKGLGANES-GIVDPVSGGEV--RNRNEQFMGLGRSLD	695 SWAP+CTBP+GP
SPF45_Hs_3746840	235 GGTVAHKIMQKYGFREGQGLGKHEQ-GLSTALSVEKT-----SKRGCKIIVGD	281 \
DRT111_At_421829	210 QMTPAQRMMPKMGWKGQGLGKSEQ-GIPTPLMAKTT-----DRRAGVIVNAS	256 GP+RRM
F58B3.7_Ce_3877880	209 GLGIAANIMKRHGYKEGAGLGKSEQ-GMSTALSIEKT-----GVRGNIVAEA	255
MHK10.5_At_4567305	103 SGGIGMKLLEKMGYK-GRGLGKNQO-GIVAPIEVQLR-----PKNMGMGYNDF	148 \
ITBA4_Hs_2274972	99 EDNIGNQLLRKMGWT-GGGLGKSGE-GIREPISVKEQ-----HKREGLGLDVE	144 /
C07E3.1b_Ce_3874094	86 NSNKIMKMMQAMGYKPGEGGLGAQQG-GIVEPVQAQLR-----KGRGAVGAYGK	132 \
49C10.14_Ce_3859703	83 KYGIGAKLLLMKMGYQKGLGVNQE-GIINPIETKLR-----PKGLGVGAVGK	129
HS1048E9A_Hs_4886427	149 TKGIGQKLLQKMGYVPGRLGKNAQ-GIINPIEAKQR-----KKGAVGAYGS	195 GP+Coiled coil +X
GH11690_Dm_5052630	167 TRGIGAKLLLQMGYEPGKGLGKDLQ-GISHPVQAHVR-----KGRGAIGAYGP	213
YLR424w_Sc_664877	61 TYGIGAKLLSMSGYVAGKGLGKDGSG-GITTPLETQSR-----PMHNAAGLGMFSN	108 /
Bat-4_Mm_1196644	271 TSSPGFRLLLLRGGWEPGGLGPRGE-GRANPIPTILK-----RDQEGGLGYRSP	317 2*ANK+GP
ZK1320.7_Ce_3881606	942 PSNIGYRLMCASGEWTEQGLGRNSD-QORFPIRTILK-----RNRAGLGAEKL	988
SON_Hs_586013	1298 TGGMGAVLMRKMGWREGEGLGKNKE-GNKEPILVDFK-----TDRKGLVAVGE	1344 \
D1037.1_Ce_3893874	637 TGGIGRLMLEKMGWRPGEGLGKDAT-GNLEPLMLDVK-----SDRKGLIAEEE	683 GP+dsRBD
YLR271w_Sc_577172	41 IMPRGYKMMENMGYKEGETLGSNES-ALKEPIKVEIN-----TKRRGIRAEPK	87
T04H1.5_Ce_3879426	73 HESKGFALMAKMGFKPGMTLQKQRE3RITFPIGVDIK-----ANRKLGHVEVE	122 \
VPRT_Hs_139377	289 YSPTSQKIMTKMGYIPGKGLGKNEDE-GIKVPVEAKIN-----QEREGIGYPF-	334 \
VPRT_MPMV_139379	270 PNDIVTAQMLAQGYSPGKGLGKKNEN-GILHIPNQGQ-----SNKKGFGNF--	314 dUTPase+Protease+GP
VPRT_OPAV_279569	281 PSPPTVTDLMLDQGLLPNQGGLGKQHQ-GIILPLDLKPN-----QDRKGLGCF--	325 /
YNL224c_Sc_2131930	720 NENIGRRMLEKLGWKSSEGLGIQGNKGISEPIFAKIK-----KNRSGLRHSES	767 /
BC30B4.02c_Sp_3417428	652 KENPGRRLLLEKLGWYAGKGLGHPENEGSKDSLRAIVK-----VSRSLG----	695 /
YGR280c_Sc_2131718	25 TSFRGHQFLEKFCWKPQMGGLGSPM-NSNTSHIKVSI-----KDDNVGLGAKLK	72 \
T23G7.3_Ce_3880143	26 DQKLSKLLMEKMGWSEGDGLGRNRQ-GNADSVLKAN-----TSGRGLGADKM	72 /
C33H5.17_Ce_1086596	248 GGGVGMKLLMMKMGYKVGEGLEKRSDE-GIVHAIQARIC-----KKNASLDEVMT	294 \
F27C12.25_At_4559368	300 TRGIA SKMMASMGYREGMGLGVSQG-GILNPIILVKVL-----PAKRSLDYALE	346 /
CC1442.13c_Sp_3790258	145 NNGKQKQLLEMMGWSRQGLGSENG-QMVDPVVAVVK-----NNKQGLH----	187
AC2G11.04_Sp_2130390	252 KENVALKMLQRCGWKEGQGLGQHNQ-GIINPLHVEIS-----GFVTETKHSKI	298
K07B1.7_Ce_2088823	396 QKGIGSMLAKMGWKPGETGLCKNEQ-GRVVPVAVYVEEDGQSSKEKTFGFGYRGE	448
R07E5.1_Ce_630724	145 SNSIGVRMLRSMGWREGRGIGLANV-KQKQKRGGESS-----EAQFDREQASK	191
F21E10.7_At_3047080	52 LEDVEQASLDMKMGW-GKGLGKQEQ-GITEPIKSGIR-----DRRLGLGKQEE	97
F13M23.160_At_4455245	139 VDFGFEALLAGYGWKPQGGIGLKAK-EDVKIVEYEKWK-----SGNEGFGFGKS	185
1484m_Pf_tr2320595	??? KKDFA TRMMEKMGWKKGEGLGKDKQ-GIKAPLILKKV-----DKRSGVIVQAP	???
SSE261_Dd_tr3984514	??? STNIGSSLLKMGWS-GGGLGVNGQ-GLEEPKAVVK-----KERSGLGFA--	???
CHR5Ex_At_tr2351062	??? SSNIGFQLLKKHGKWTGLGIAEQ-GILVPLQAEPK-----HNKRGVGAKQP	???
consensus/90%	..s.s..hh..Ga..G.GlG...p.u...s.....u.....	



TIBS 24:
342-344
(1999)

RNA-binding
Domain

Introduction to OCRE Domain

		1	2	3	4	5		
VG5Q	Hs	AEEAAVSQTGFSYDEN	---TGLYFDH-STGFYYDSE	---NQLYYDPSTG	---IYYYCDVESG	---RYQFHSRV	197-256	45708564
CNBE2600	Cn	MTSGAAPQDTWVFDSS	---TGIFYHAPSNTYATHDSATGQWYHTASE	---FHQASGSNI	---HHQQSLED		1-62	50258205
ENSANGP00000013789	Ag	TAQDVDIQNRYYEPT	---SGMYYP-ATGYYYNSI	---YGLHYDGHGRG	---CYLKYNEETK	---EYDFYSQV	48-107	31217017
4N366	Ce	EVLQSRIPPGYVLYPN	---YEQYWSH-EHGMFYAPD	---TEQFYRPEDK	---TVYQLNAEGS	---EYVLVERR	14-73	17543636
At3g54230	At	QKGTSAPOSGYVWDEA	---SGYYYDA-ASGYYYDGN	---SGLYYDSNSG	---LWYSYDQQTQ	---QYVPCPDQ	574-633	17979131
RBM5	Hs	TKYAVPDTSTYQYDES	---SGYYYDP-TTGLYYDPN	---SQYYNSLTQ	---QYLYWDGEKE	---TYVPAAES	452-511	5032031
AT02677p	Dm	KKYDTPDVSKYQYDET	---SGYYYDH-VTGLYYDAH	---SQYYNNETG	---AYLYWDQKRS	---TYVLATPA	443-502	25009684
RBM10	Hs	SQYPVPDVSTYQYDET	---SGYYYDP-QTGLYYDPN	---SQYYNAQSQ	---QYLYWDGERR	---TYVPALEQ	561-620	12644371
CG4887-PA	Dm	KIYPTPDVSTYQYDET	---SGYYYDS-TTGLYYNAH	---SQYYNNETG	---AYLYWDQRRS	---TYVLATPA	620-679	24580815
MGC68576	Xl	LKYSVPDSTYQYDES	---SGYYYDP-QTGLYYDPN	---SQYYNSLTQ	---QYLYWDGEKQ	---TYLPAADG	325-384	37589344
CG8079-PA	Dm	AAKHAENLNFFVYEHT	---SGMYYP-KTGYYYNAE	---YGLYYDGNNG	---CYYSYDHAKD	---SYEFHSQA	121-180	19922310
1G795	Ce	PRYLPPNPLTFADPN	---LGYYVDP-ITKFCYDSA	---TGYYFNATS	---QWCNWDLTHH	---TYFPVETP	463-522	32563880
P0583G08.4	Os	AVDGEASGAREWDED	---SGLYYHA-STGFYHPV	---AGWYSSGRDG	---QYYIYDNG	---NYVPWTPD	120-177	15528704
RBM6	Hs	DHMHYQGGKRYFRORR (21)	DCYIYDS-ATGYYYDPL	---AGTYYPNTQQEVYVPQDPGLPEEEEEIKKKPT			751-836	12643357
At1g49590	At	NGAPEDGTSDWMLDSA	---SGYYYNQ-TNGLHYDSQ	---SGFYSDSIG	---HWVTQDEAYA	---AVKTSSTG	90-149	15222782
F28A23.100	At	QFLKPKFEKRLLIESS (3)	SGFYHDP-NAGWYYCSK	---DGRYKHEG	---EYVPLEYDES	---SNPPGDI	19-81	7460046
At5g09390	At	SISGTQLNSDYVDES	---SGYYYSS-SLGYYPN	---TGLYCYATTG	---KWYRYDEETK	---EYKEVNSE	286-345	18415962
PF13_0278	Pf	LKKGKPELSEFFFDYN	---SQYYYHP-LYQLYFDNN	---TKYMSLSKG	---YYIWEDELK	---CLLRVYLD	796-854	23619456
PY04659	Py	LKNEKPOLTEFFFDYN	---SKYYYHP-TYQIYFDNN	---TSFYMSLANG	---YYIWNENSK	---CLVRFYTD	182-240	23479825

SS prediction EEEE EEE EEE EEE EEEEE EE

Affiliations, *Bioinformatics* 21: 699-702 (2005)

Phe, Tyr, Val, Leu, Ile, Thr rich, 5-fold, imperfectly repeated octameric sequence.

Functional unknown, often associated with RNA binding domain.



Thanks!

