

Applied Bioinformatics Course

拟南芥中的At3g03950生物信息学分析 Bioinformatics analysis of At3g03950 in *Arabidopsis Thaliana*

Group : G11
魏连环 张笑 张健
20150627



概要

1 背景

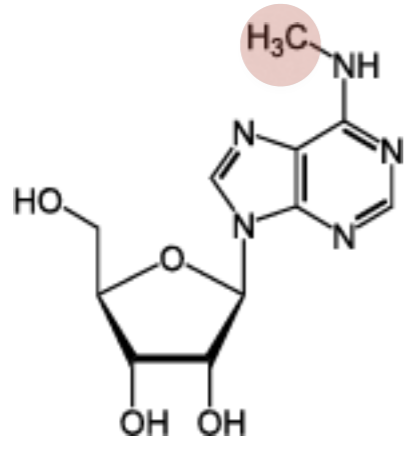
2 生物信息学分析

- 序列分析
- 结构分析
- 功能分析

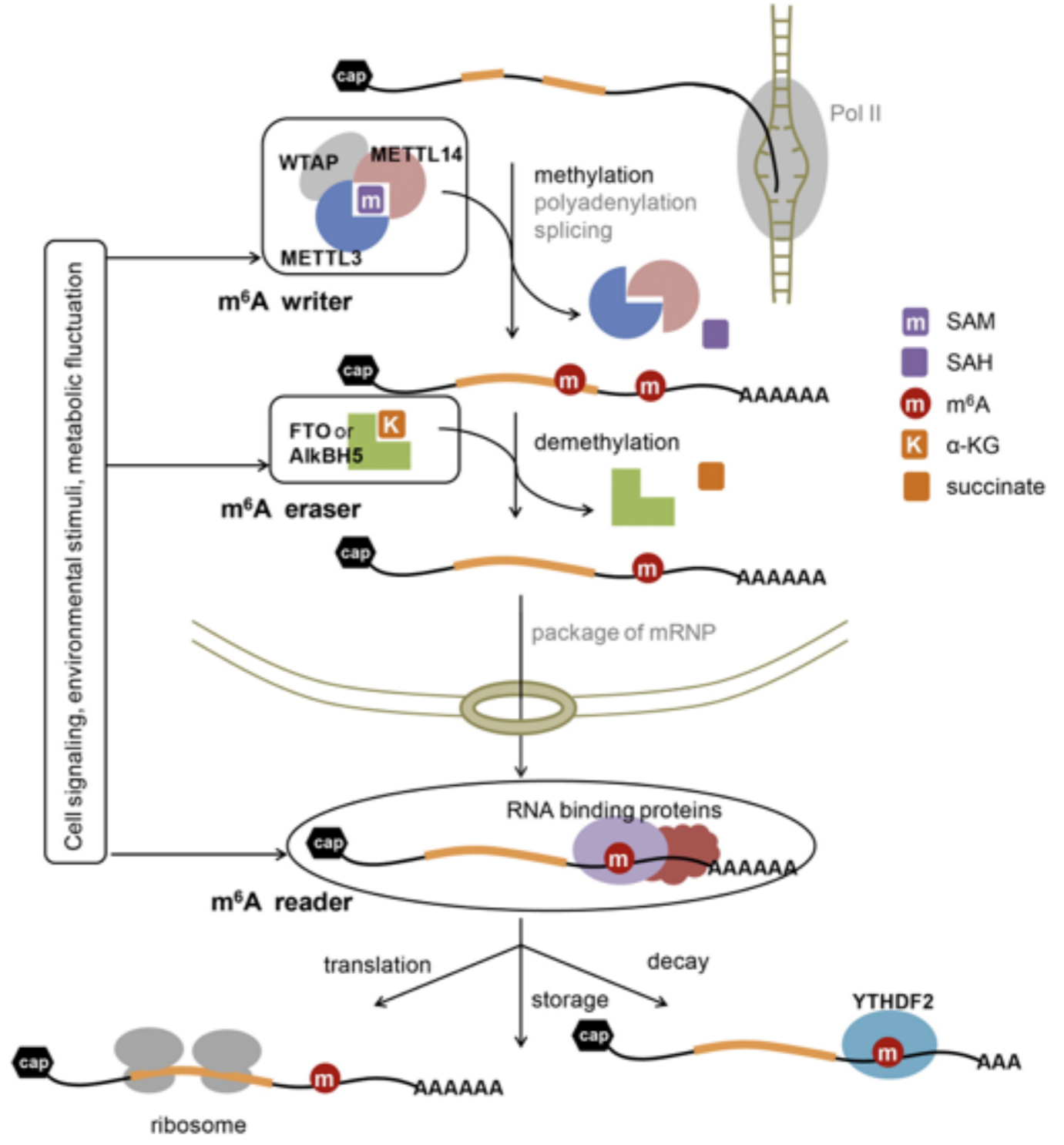
3 总结

4 致谢

背景



N^6 -Methyladenosine (m^6A)



Wang, X.; He, C., Dynamic RNA modifications in posttranscriptional regulation. *Mol Cell* **2014**, *56* (1), 5-12.

拟南芥中YTH结构域广泛存在

序列: YTHD2_HUMAN(410-544) 程序:blastp 参数:默认 数据库:Reference Protein 物种:Arabidopsis thaliana Query: yth

	Description	Max score	Total score	Query cover	E value	Ident	Accession	
Select seq ref	NP_565205.1 uncharacterized protein [Arabidopsis thaliana]	216	216	53%	4.00E-65	39%	NP_565205.1	At1g79270 Y6 528 aa
Select seq ref	NP_187912.2 protein ECT5 [Arabidopsis thaliana]	218	218	30%	8.00E-65	56%	NP_187912.2	At3g13060 Y8 634 aa
Select seq ref	NP_851236.1 evolutionarily conserved C- terminal region 3 protein [Arabidopsis thaliana]	208	208	38%	1.00E-62	46%	NP_851236.1	At5g61020 Y13 495 aa
Select seq ref	NP_187955.2 CIPK1 interacting protein ECT2 [Arabidopsis thaliana]	211	231	34%	2.00E-62	54%	NP_187955.2	At3g13460 Y9 667 aa
Select seq ref	NP_974954.1 evolutionarily conserved C- terminal region 10 [Arabidopsis thaliana]	208	208	35%	3.00E-62	47%	NP_974954.1	At5g58190 Y12 528 aa
Select seq ref	NP_001185240.1 uncharacterized protein [Arabidopsis thaliana]	209	209	28%	5.00E-62	58%	NP_001185240.1	At1g55500 Y5 599 aa
Select seq ref	NP_174117.2 uncharacterized protein [Arabidopsis thaliana]	202	225	31%	4.00E-60	58%	NP_174117.2	At1g27960 Y2 539 aa
Select seq ref	NP_175245.1 uncharacterized protein [Arabidopsis thaliana]	203	203	29%	1.00E-59	52%	NP_175245.1	At1g48110 Y4 639 aa
Select seq ref	NP_001030629.1 CIPK1 interacting protein [Arabidopsis thaliana]	195	195	31%	2.00E-58	51%	NP_001030629.1	At3g03950 Y7 428 aa
Select seq ref	NP_188359.2 evolutionarily conserved C- terminal region 6 [Arabidopsis thaliana]	199	220	38%	2.00E-58	50%	NP_188359.2	At3g17330 Y10 595 aa
Select seq ref	NP_172452.3 evolutionarily conserved C- terminal region 11 protein [Arabidopsis thaliana]	184	184	30%	5.00E-54	47%	NP_172452.3	At1g09810 Y1 470 aa
Select seq ref	NP_192934.2 YTH family protein [Arabidopsis thaliana]	77.8	77.8	26%	5.00E-18	32%	NP_192934.2	At4g11970 Y11 444 aa
Select seq ref	NP_174334.2 cleavage and polyadenylation specificity factor CPSF30 [Arabidopsis thaliana]	75.9	75.9	27%	3.00E-17	27%	NP_174334.2	At1g30460 Y3 631 aa

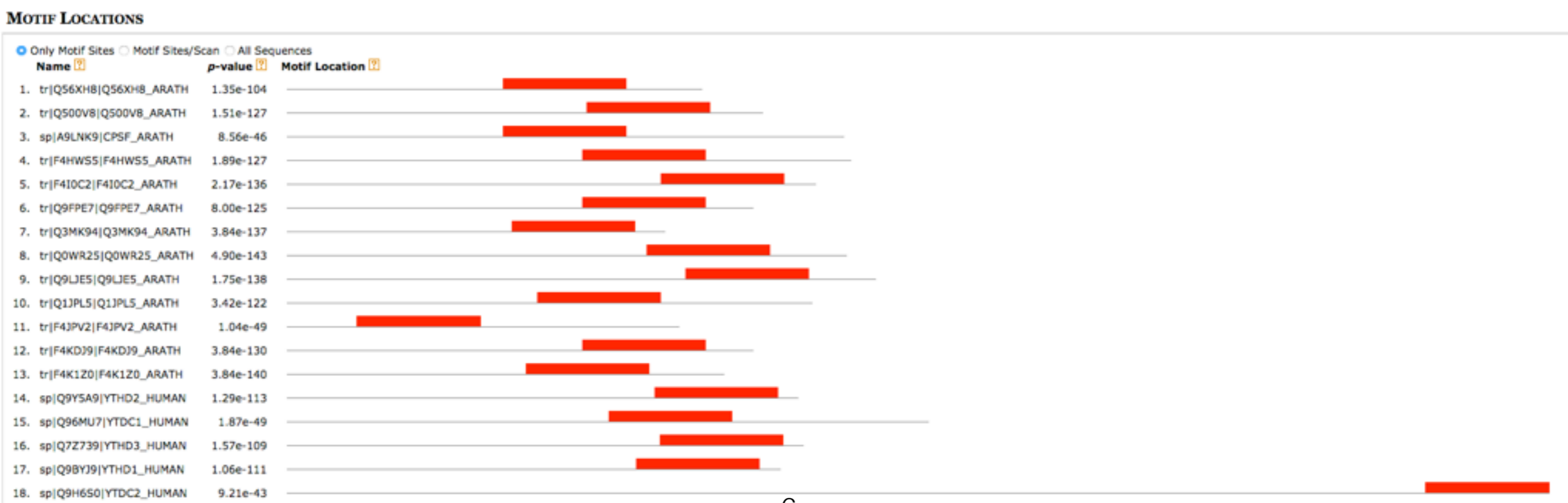
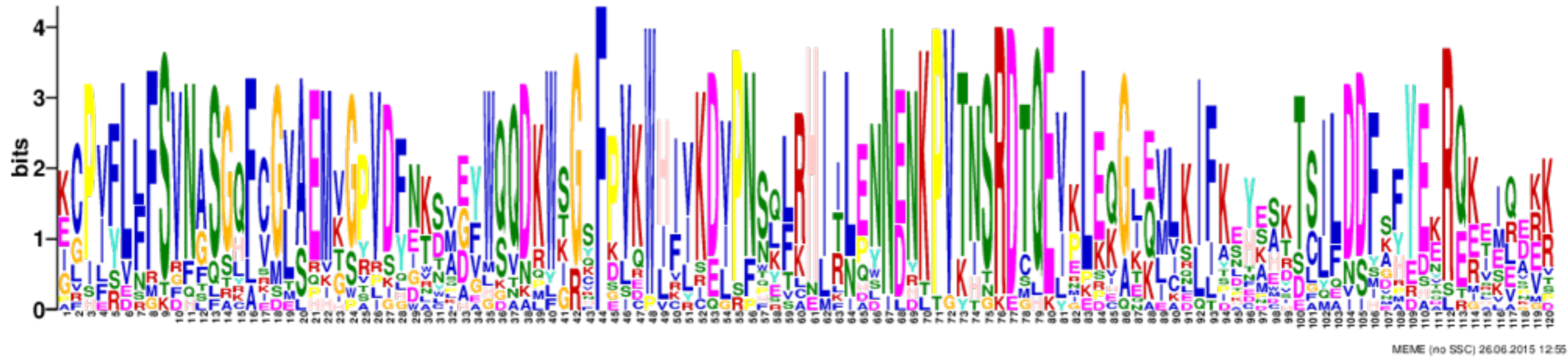
序列分析

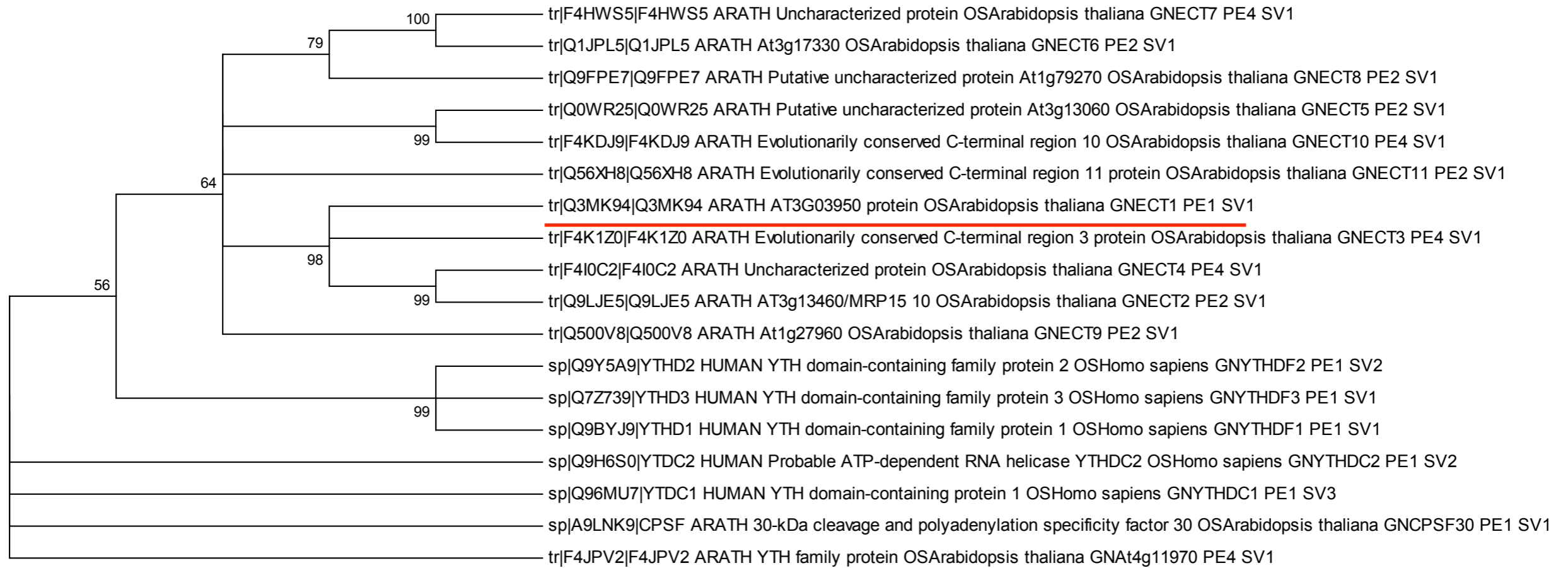
YTH结构域非常保守

Q56XH8	Q56XH8_ARATH	203	-----PLKQNNSPALALRREMYNLPDFQTDYEDAKF	FVVIKSYSEDDVHKS	SIKYSVWSSTI	257										
Q500V8	Q500V8_ARATH	312	-----YDRVDRFCQOELL	SQFRDAKF	FVVIKSYSEDDVHKS	SIKHCWASTK	356									
A9LNK9	CPSF_ARATH	208	Q--SQHQVSQTLIPN--PADQTNRTSHPLPQGVNRY	FVVKSNREN	FELSVQQGVWATQR	263										
F4HWS5	F4HWS5_ARATH	291	---TKAGNADAEGNIVINPSQYNKEDLRIDYSNAKF	FVVIKSYSEDDVHKS	SIKYNVWSSTL	347										
F4I0C2	F4I0C2_ARATH	384	---DESNTTEETVTCVLPDREECNRDDFPVEYKDAKF	FVVIKSYSEDDVHKS	SIKYNVWASTP	440										
Q9FPE7	Q9FPE7_ARATH	288	SLDAEGNERSNGVGSVIRRDQYNLPSFQTKYEEAIF	FVVIKSYSEDDIHK	SIKYNVWSSTL	347										
Q3MK94	Q3MK94_ARATH	216	S----MLDAMKQDVSAVDLQRYNGENFPESFVKAKF	FVVIKSYSEDDVHNC	IKYGAWSSTP	271										
Q0WR25	Q0WR25_ARATH	370	-----ADSKKNNKGS	AKEHEESNADFVTDY	TNAKLF	FVVIKSYSEDDVHKS	SIKYNVWASTP	424								
Q9LJE5	Q9LJE5_ARATH	409	ESNVTEVGEADNTCVVPDREQYNKEDFPVDYANAMF	FVVIKSYSEDDVHKS	SIKYNVWASTP	468										
Q1JPL5	Q1JPL5_ARATH	240	---TKAGNADAEGNIVINPDRYNKEDFSIEYS	DARF	FVVIKSYSEDDVHKS	SIKYGWSSSTL	296									
F4JPV2	F4JPV2_ARATH	43	L--SKVDVDRRNFDP--QLES	AKANNSKPGYRTRY	FVVIKSLNYDNI	QVSVEKGIWATQV	98									
F4KDJ9	F4KDJ9_ARATH	294	--STIGDSASDSSTAGPNPSLYNHPEFVTDYKNAKF	FVIVKSFSEDDNVHRS	SIKYNVWASTP	351										
F4K120	F4K120_ARATH	232	T----ETEKLSEDVSLDDPKDYNKIDFPET	YTEAKF	FVVIKSYSEDDIHK	SIKYSVWSSTP	287									
Q9Y5A9	YTHD2_HUMAN	382	-----PSEHPVLEK	LRSINNYNPKDFDWN	LKHGRV	FVVIKSYSEDDIHR	SIKYNICWSTE	436								
Q96MU7	YTDC1_HUMAN	330	-----HEKLS	SVRAV--RKDQ--	TSKLYVLQDARF	FLIKSNHENV	SLAKAKGVWSTLP	381								
Q7Z739	YTHD3_HUMAN	388	-----SVEVHPVLEK	LKAANNYNPKDFDWN	LKNGRV	FVVIKSYSEDDIHR	SIKYSIWCSTE	442								
Q9BYJ9	YTHD1_HUMAN	361	-----SVESH	PVLEK	LKAHSYNPKFEWNLK	SGRV	FVVIKSYSEDDIHR	SIKYSIWCSTE	415							
Q9H6S0	YTDC2_HUMAN	1257	D--SSSY	SPCASPPSSGK	SKSPSPRP	NMPVRY	FIMKSSNLRN	LEISQQKGIWSTTP	1314							
					:::***	*.:							
Q56XH8	Q56XH8_ARATH	258	NGNKKLDAAFRDAETKTLEDGKKRP	IFLFFSVNAS	RQFVGLAEMV	GYVDFN	KDLDFWQ--	315								
Q500V8	Q500V8_ARATH	357	NGNKKLDAAYREAKK----	KDVACP	VFLFLFSVNAS	SQFCGVAEMV	GPVDFNTSVEY	WQ--	410							
A9LNK9	CPSF_ARATH	264	SNEAKLNEAFDSVEN-----	VILIFSVNR	TRHFQGC	AKMTRS	RIGGYIGGGN	WKHE	313							
F4HWS5	F4HWS5_ARATH	348	HGNKKLQSAVEDAQRIATEKSCECP	IFLFFSVNAS	GLFCGMAEMT	GPVSF	DKDMDFWQ--	405								
F4I0C2	F4I0C2_ARATH	441	NGNKKLDAAYQEAQQ----	KSSGCP	VFLFFSVNAS	GQFIGLAEMK	GPVDFNKNIEY	WQ--	494							
Q9FPE7	Q9FPE7_ARATH	348	NGNKKLDSAYQESQKKAADKSGKCP	VFLFFSVNAS	GQFCGVAEM	IGRV	VDYEKSME	FWQ--	405							
Q3MK94	Q3MK94_ARATH	272	TGNKKLNAAYYEAKE----	NSQEC	PVYLLFSVNAS	GQFVGLAEMV	GPVDFNKTMEY	WQ--	325							
Q0WR25	Q0WR25_ARATH	425	NGNKKLDAAYREAKD----	EKEPC	PLFLFSVNAS	SQFCGVAEMV	GPVDFEKS	VDYWQ--	478							
Q9LJE5	Q9LJE5_ARATH	469	NGNKKLAAAYQEAQQ----	KAGGCP	IFLFFSVNAS	GQFVGLAEMT	GPVDFNTNVEY	WQ--	522							
Q1JPL5	Q1JPL5_ARATH	297	NGNKKLQSVYEDAQRIATEKSRECP	IFLFFSVN	SSGLFCGVAEMT	GPVSF	DRDMDFWQ--	354								
F4JPV2	F4JPV2_ARATH	99	MNEPILEGAFHKSGR-----	VILIFSVN	MSGFFQ	GYAEMLS	PVG--	WRRDQIWSQ	147							
F4KDJ9	F4KDJ9_ARATH	352	HGNKKLDTAYRDAEK----	MGGKCP	IFLFFSVNAS	GQFCGVSE	MVGPVDFE	KDAGYWQ--	405							
F4K120	F4K120_ARATH	288	NGNKKLDASYNEAKQ----	KSDGCP	VFLFSVNTS	GQFVGLAEMV	GPVDFNKTVEY	WQ--	341							
Q9Y5A9	YTHD2_HUMAN	437	HGNKRLDAAYRSM-----	NGKGP	VYLLFSVNGS	GHFCGVAEMK	SAVDYNTCAG	VWS--	487							
Q96MU7	YTDC1_HUMAN	382	VNEKKLNLAFRSARS-----	VILIFSV	RESGK	FQGFARLS	SESHGG	SPIHWVLP	431							
Q7Z739	YTHD3_HUMAN	443	HGNKRLDAAYRSL-----	NGKGP	LYLLFSVNGS	GHFCGVAEMK	SVVDYNAYAG	VWS--	493							
Q9BYJ9	YTHD1_HUMAN	416	HGNKRLDSAFRCM-----	SSKGP	VYLLFSVNGS	GHFCGVAEMK	SPVDYGT	SAGVWS--	466							
Q9H6S0	YTDC2_HUMAN	1315	SNERKLNRAFWESSI-----	VYLV	FSVQSG	GHFQGF	SRMSSE	IGRE-KSQDWG--	1361							
			:	*	:	:	*,***:	* * :.:	.	*						
Q56XH8	Q56XH8_ARATH	316	----VDKWSG	FPVEWHVVK	DIPNWELRHI	ILDN	NEDKPV	THTRD	THEIKLKEGLQ	MLSI	371					
Q500V8	Q500V8_ARATH	411	----QDRWSG	HPVQWLIVK	DVPNSLFRHI	IESND	NKPV	TNSR	DTQEV	GLEKGIEM	LDI	466				
A9LNK9	CPSF_ARATH	314	HG--TAQYGR	NFSVKWLK	CELSFHK	TRNLRN	PNYENL	PVKISR	DCQ	ELEPSVGE	QLASL	371				
F4HWS5	F4HWS5_ARATH	406	----QDKWSG	SFPVKWHI	IKDVPNSY	FRHI	ILQNN	ENKPV	TNSR	DTQ	EVMLKQGLE	VLKI	461			
F4I0C2	F4I0C2_ARATH	495	----QDKWTG	SFPLKWHI	LKDVNSLL	KHITL	EYNN	ENKPV	TNSR	DTQ	EVKLEQGL	KVVKI	550			
Q9FPE7	Q9FPE7_ARATH	406	----QDKWTG	YFPVKWHI	IKDVPNP	QLRHI	ILENN	ENKPV	TNSR	DTQ	EVRLPQ	GNEVLNI	461			
Q3MK94	Q3MK94_ARATH	326	----QDKWIG	CFPVKWHI	IKDIPNS	LLRHI	TLANN	ENKPV	TNSR	DTQ	EVNLEH	GTKI	IKI	381		
Q0WR25	Q0WR25_ARATH	479	----QDKWSG	QFPVKWHI	IKDVPNS	QFRHI	ILENN	DNKPV	TNSR	DTQ	EVKLEQ	GIEM	LKI	534		
Q9LJE5	Q9LJE5_ARATH	523	----QDKWTG	SFPLKWHI	VKDVNSLL	KHITL	EYNN	ENKPV	TNSR	DTQ	EVKLEQ	GLKIVKI	578			
Q1JPL5	Q1JPL5_ARATH	355	----QDKWSG	SFPVKWHI	IKDVPNSY	FRHI	ILHNN	ENKPV	TNSR	DTQ	EBIILKQ	GLEVLKI	410			
F4JPV2	F4JPV2_ARATH	148	GGK--NNPWGR	SFKVWLRL	SELPP	FQKTLHL	KNPL	NDYK	PVKISR	DCQ	ELPEDI	GEAL	CEL	206		
F4KDJ9	F4KDJ9_ARATH	406	----QDRWSG	QFPVKWHI	VKDIPN	NRFC	HILLQ	NNDNKPV	THSR	DSQ	EVKLRQ	GIEM	LRI	461		
F4K120	F4K120_ARATH	342	----QDKWIG	CFPVKWHI	VKDIPNS	SLRHI	TLENN	ENKPV	TNSR	DTQ	EVKLEQ	GIKVIKI	397			
Q9Y5A9	YTHD2_HUMAN	488	----QDKWKG	RFVVRWIF	VKDVPNS	QLRHI	RLENN	ENKPV	TNSR	DTQ	EVPLEK	AKQVLKI	543			
Q96MU7	YTDC1_HUMAN	432	AGMSAK	MLGGVFKID	WICRREL	PFTK	SAHL	TNPW	NEHK	PVKIG	RDGO	EBIE	LEC	GTQL	CLL	491
Q7Z739	YTHD3_HUMAN	494	----QDKWKG	KFEVKWI	FVKDVP	NNQLRHI	RLENN	DNKPV	TNSR	DTQ	EVPLEK	AKQVLKI	549			
Q9BYJ9	YTHD1_HUMAN	467	----QDKWKG	KFDVQWIF	VKDVP	NNQLRHI	RLENN	DNKPV	TNSR	DTQ	EVPLEK	AKQVLKI	522			
Q9H6S0	YTDC2_HUMAN	1362	----SAGLGG	VEKVEWIK	ESLPP	FQFAHLL	NPW	NDNKKV	QISR	DCQ	EBLE	PLVGE	QLLQL	1417		
			*	:	*	:	:	*	:	*	**	:.:	.	:	:	

序列分析

YTH结构域非常保守





基本信息

Entry name: Q3MK94_ARATH

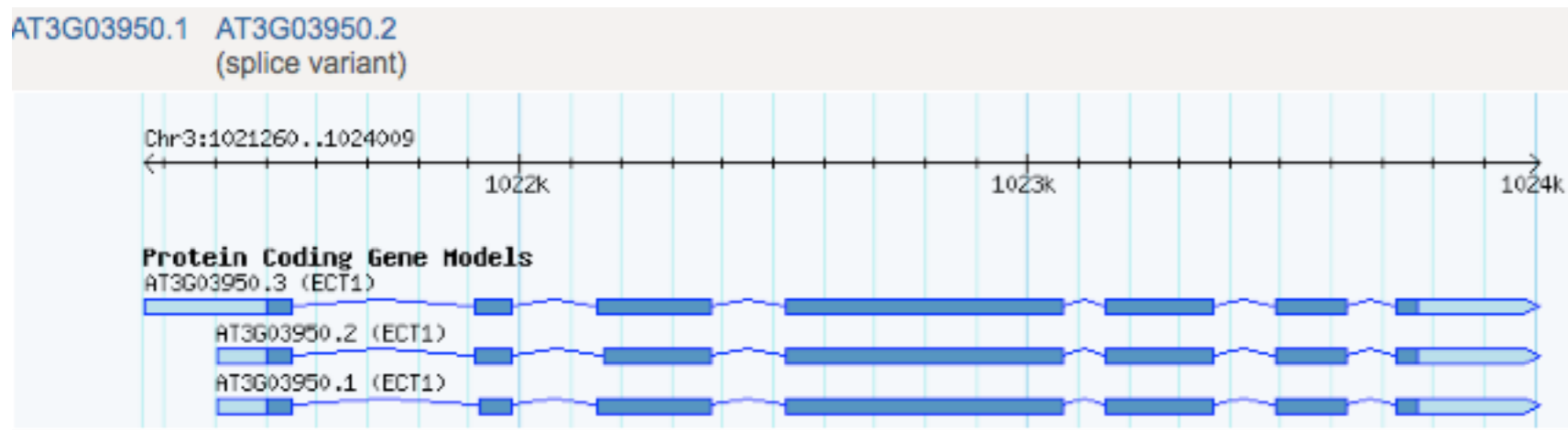
Submitted name: AT3G03950 protein

Gene: ECT1

Organism: Arabidopsis thaliana (Mouse-ear cress)

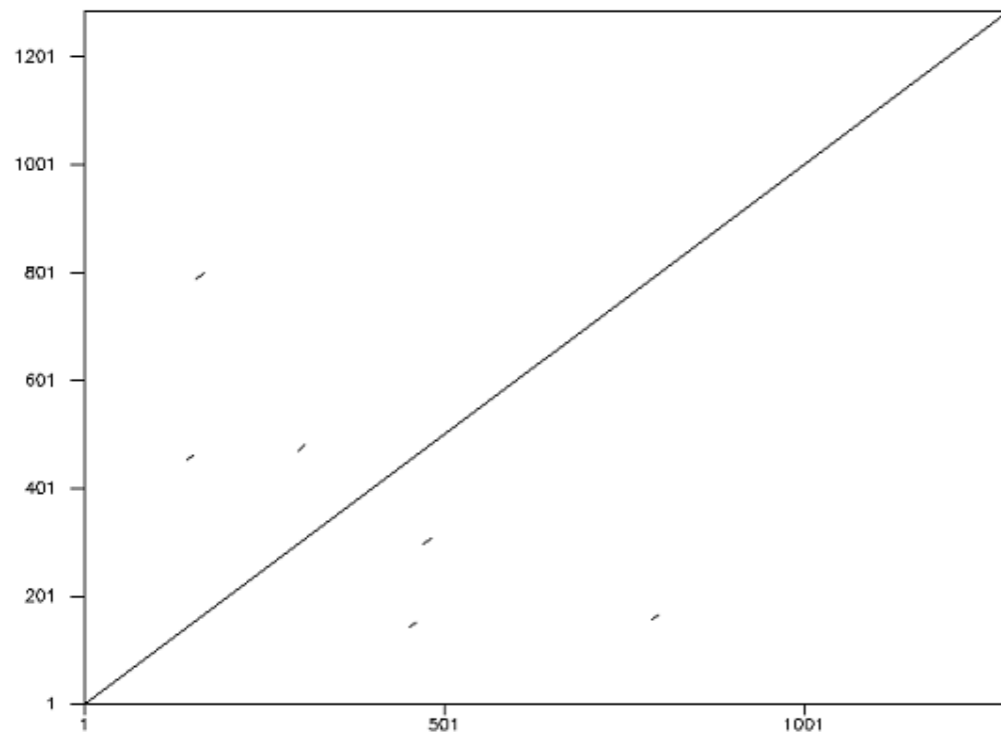
Status: Unreviewed

Experimental evidence at protein level

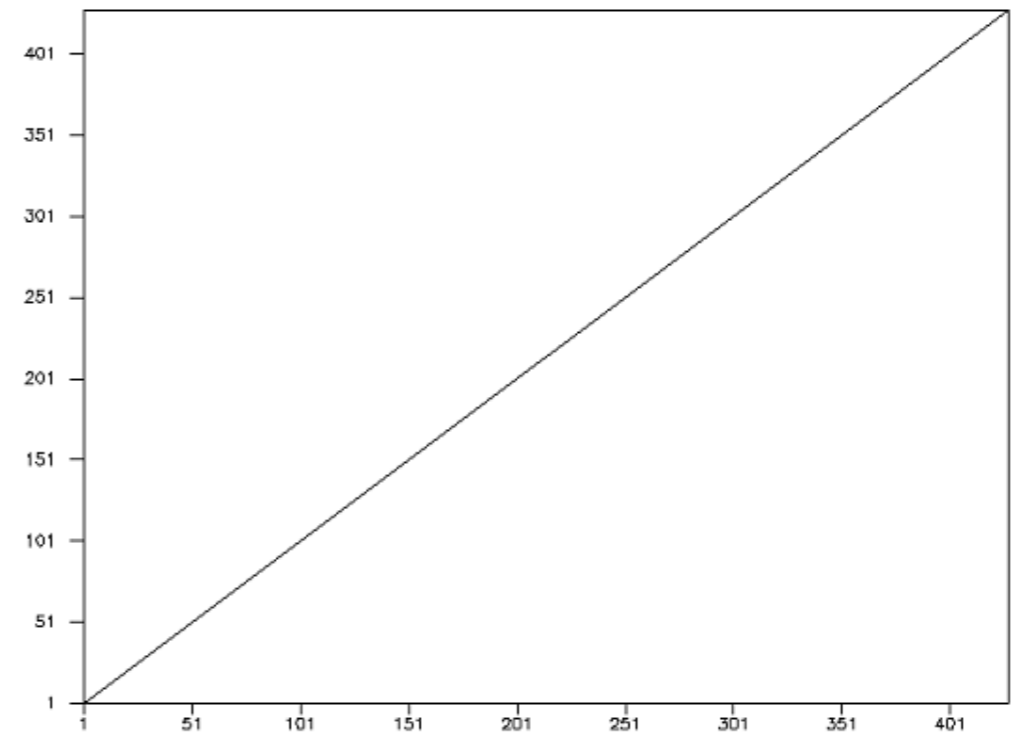


重复序列预测

Dottup_CDS

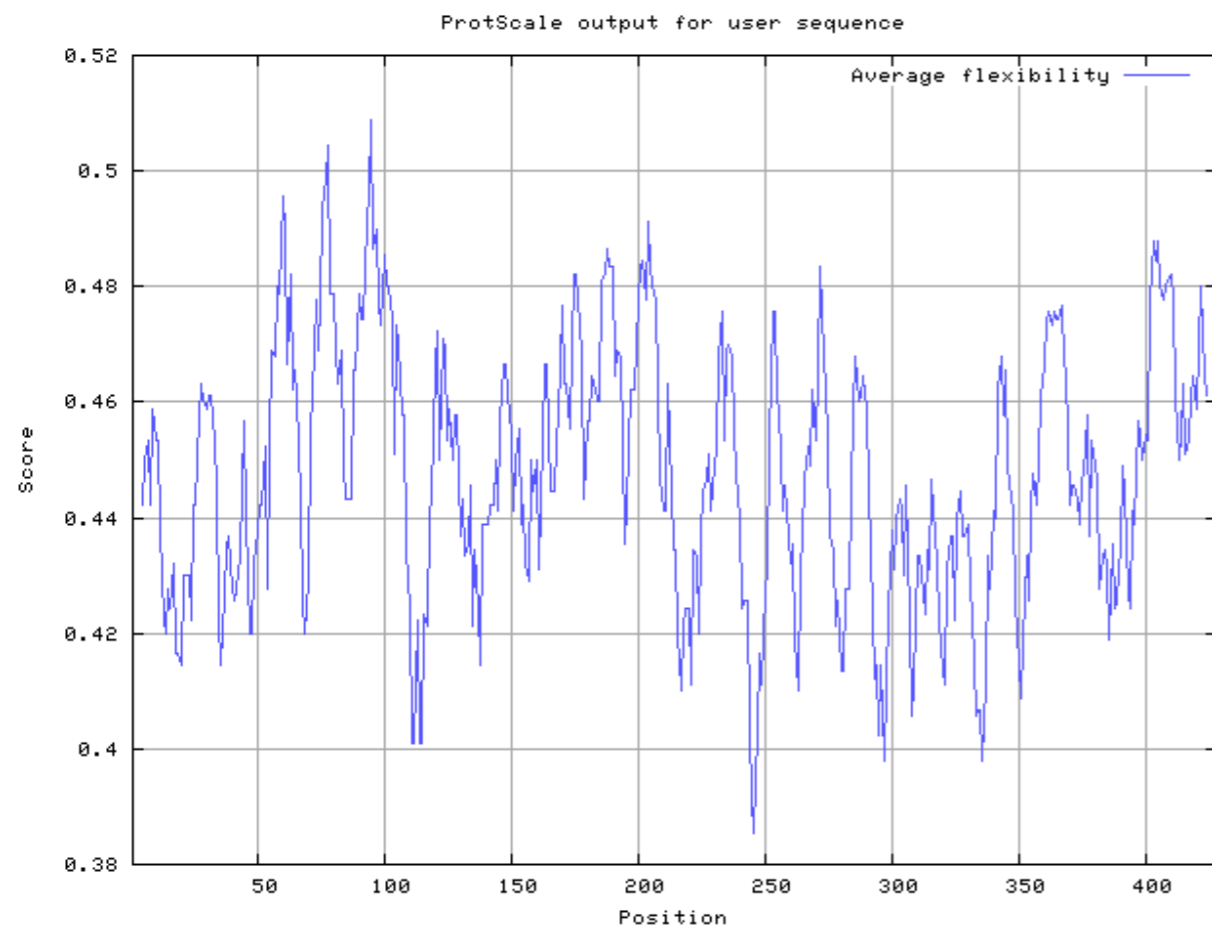


Dottup_AA

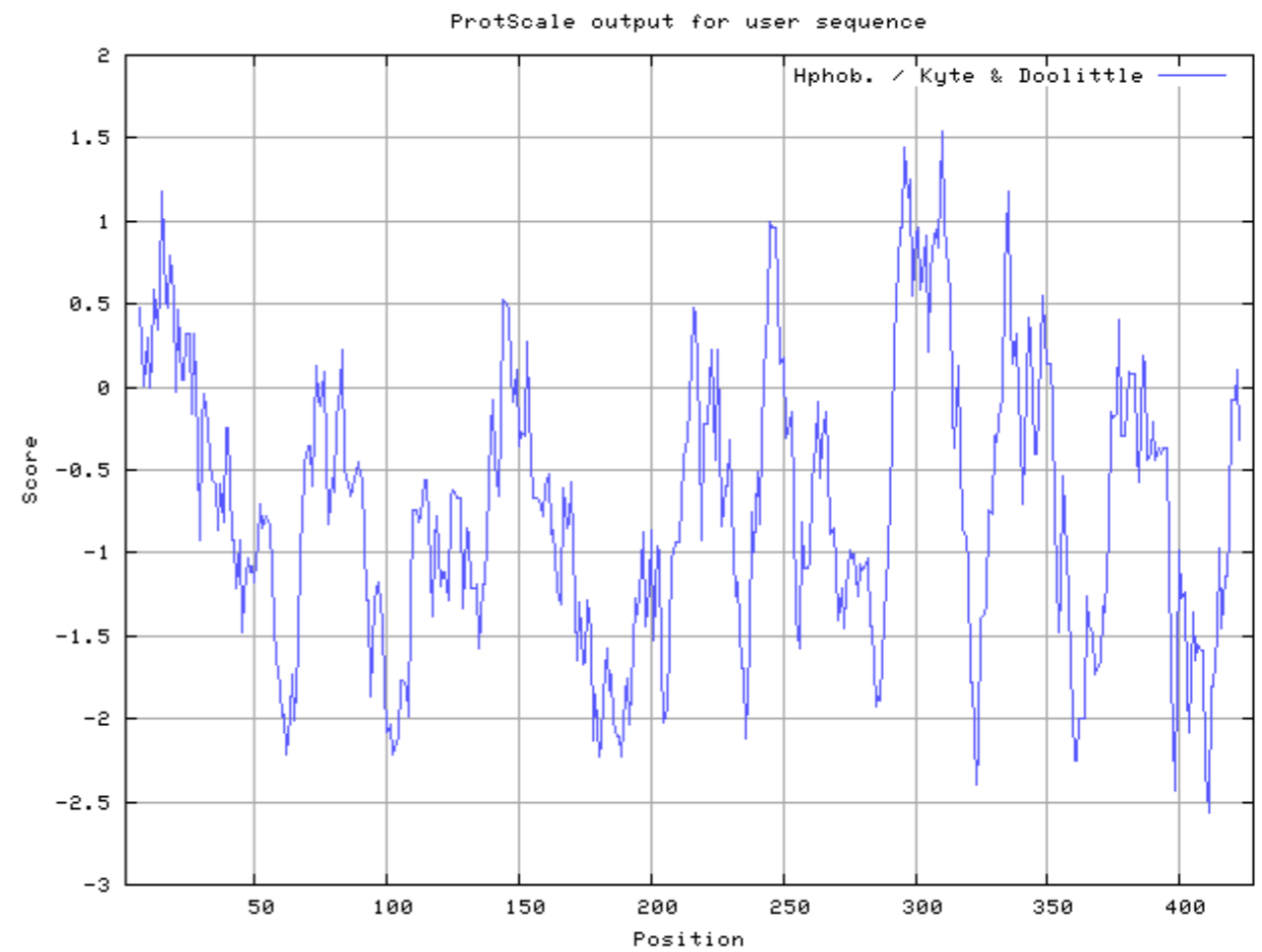


化学性质分析

Average flexibility₁



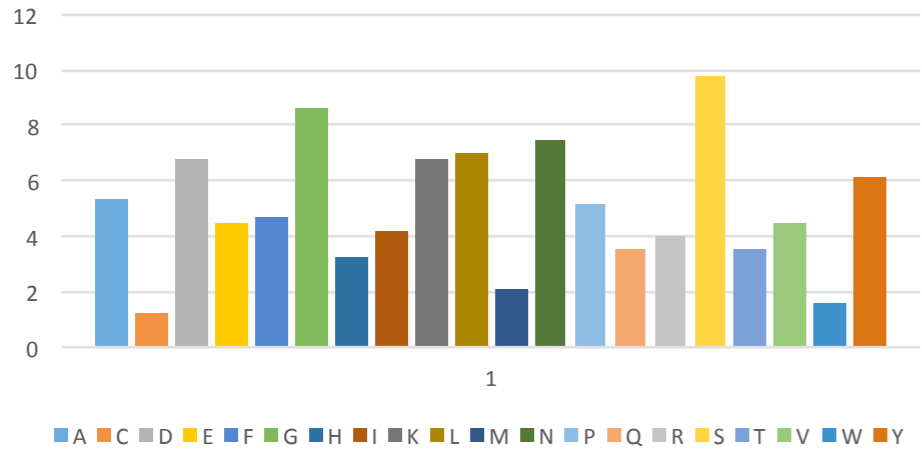
Hphob. / Kyte & Doolittle₁



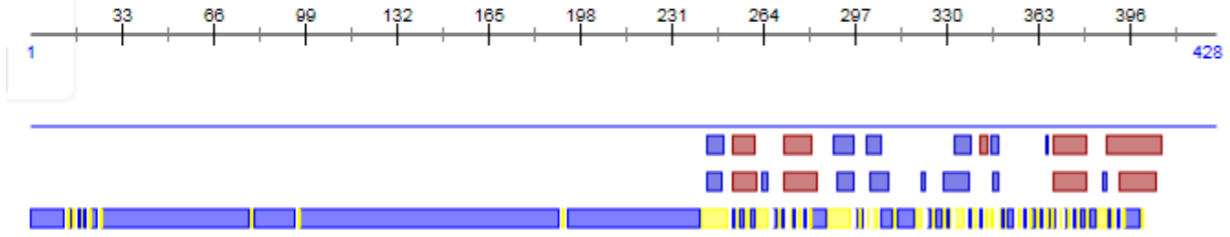
结构分析

ProtParam

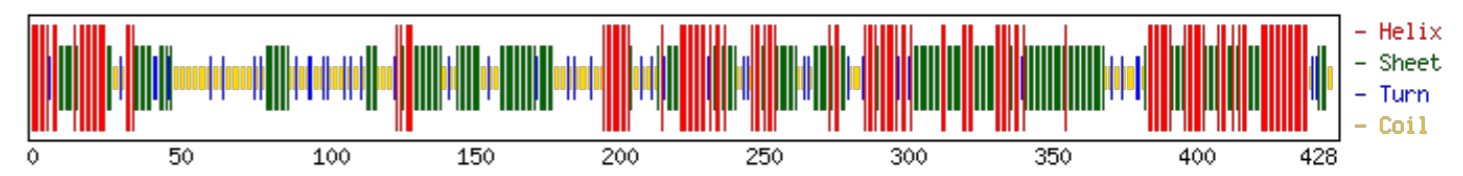
氨基酸组成



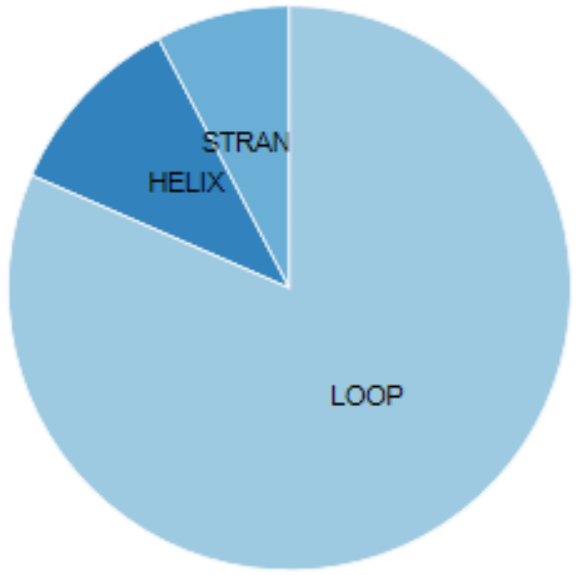
A C D E F G H I K L M N P Q R S T V W Y



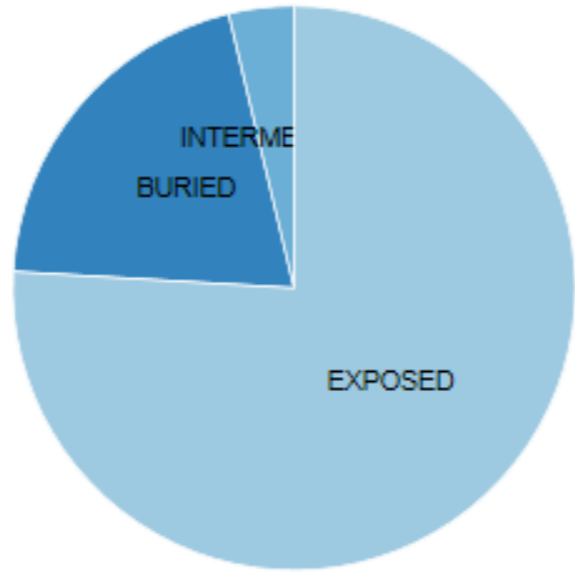
CFSSP: Chou & Fasman Secondary Structure Prediction Server

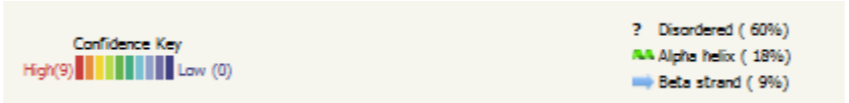


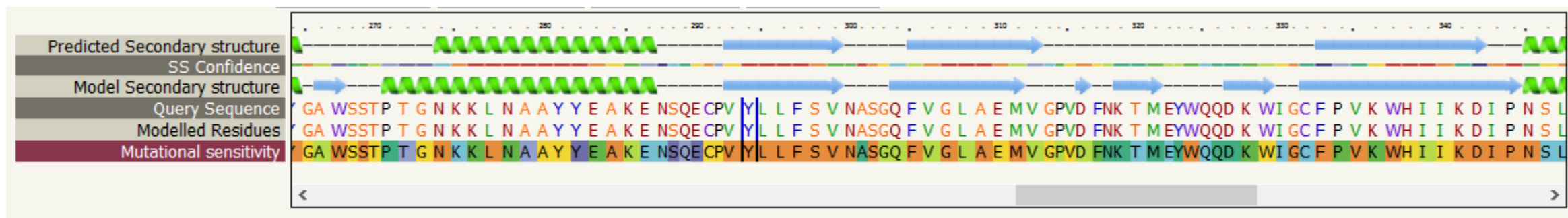
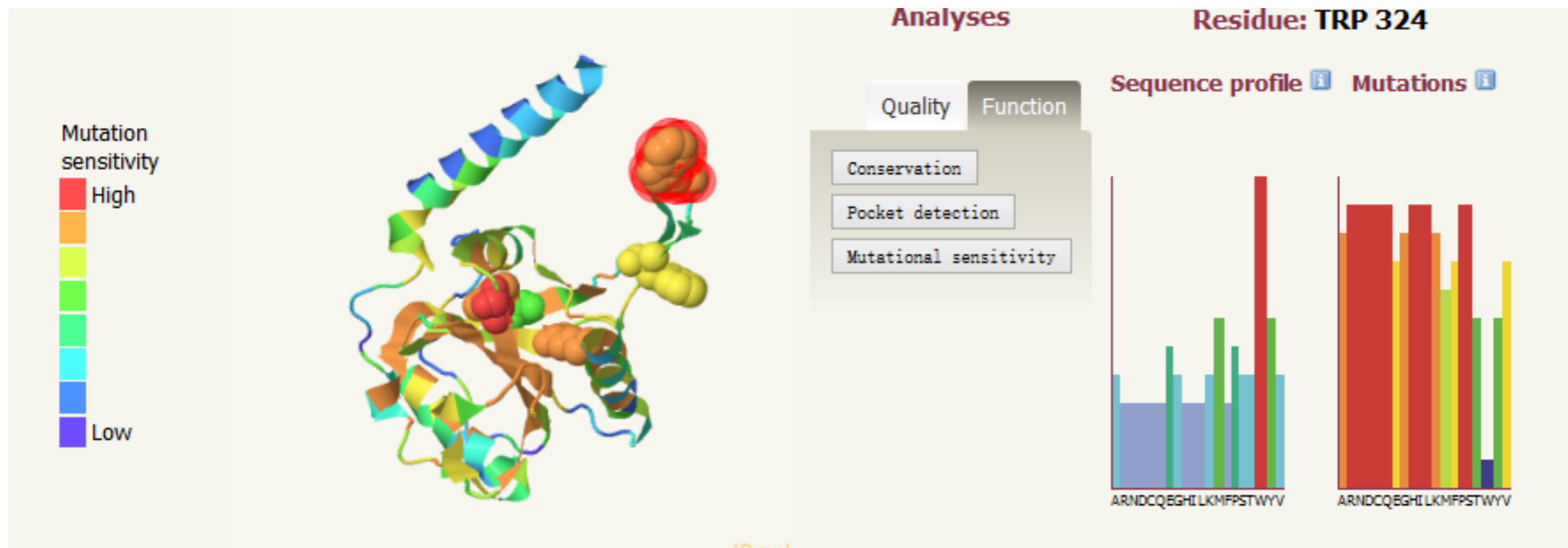
Secondary Structure Composition

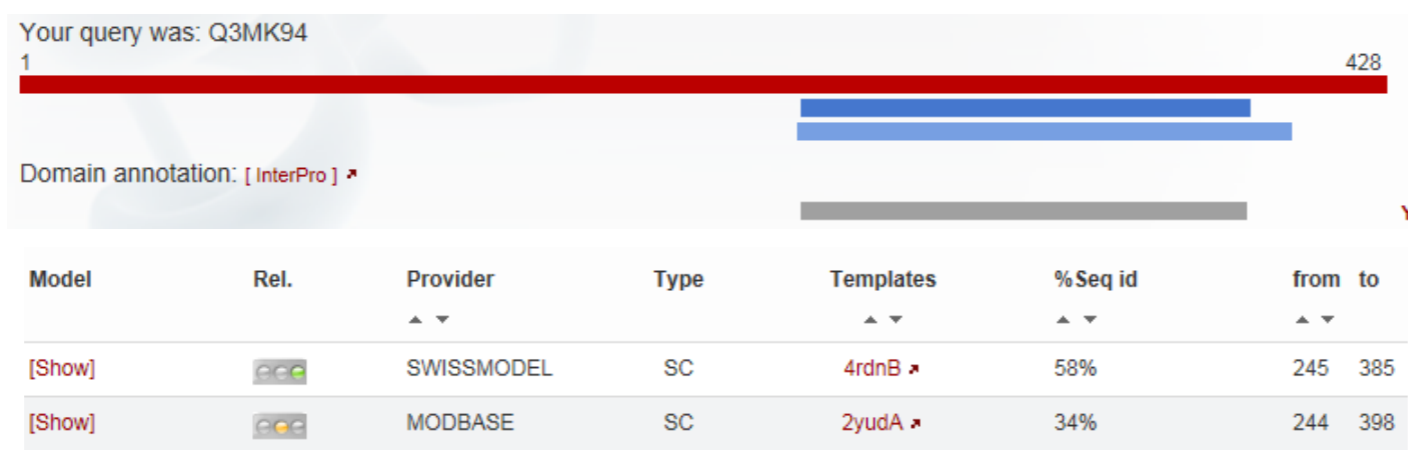


Solvent Accessibility



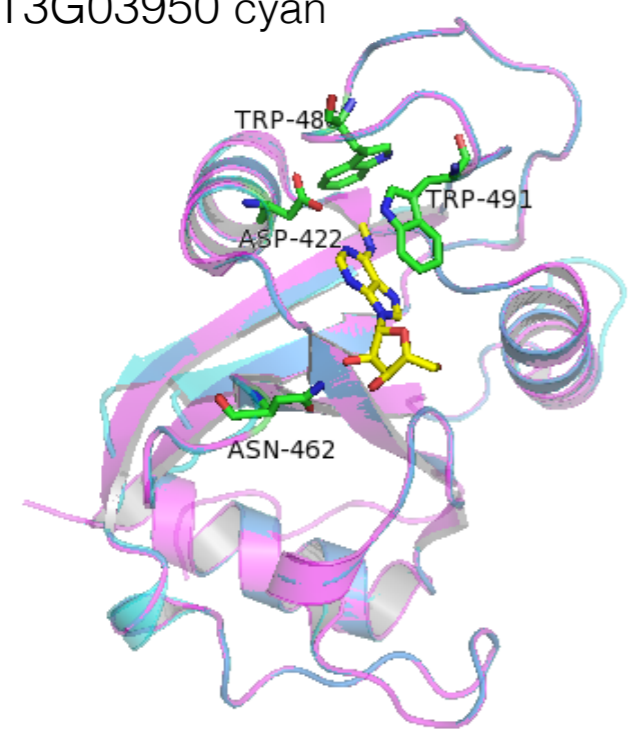




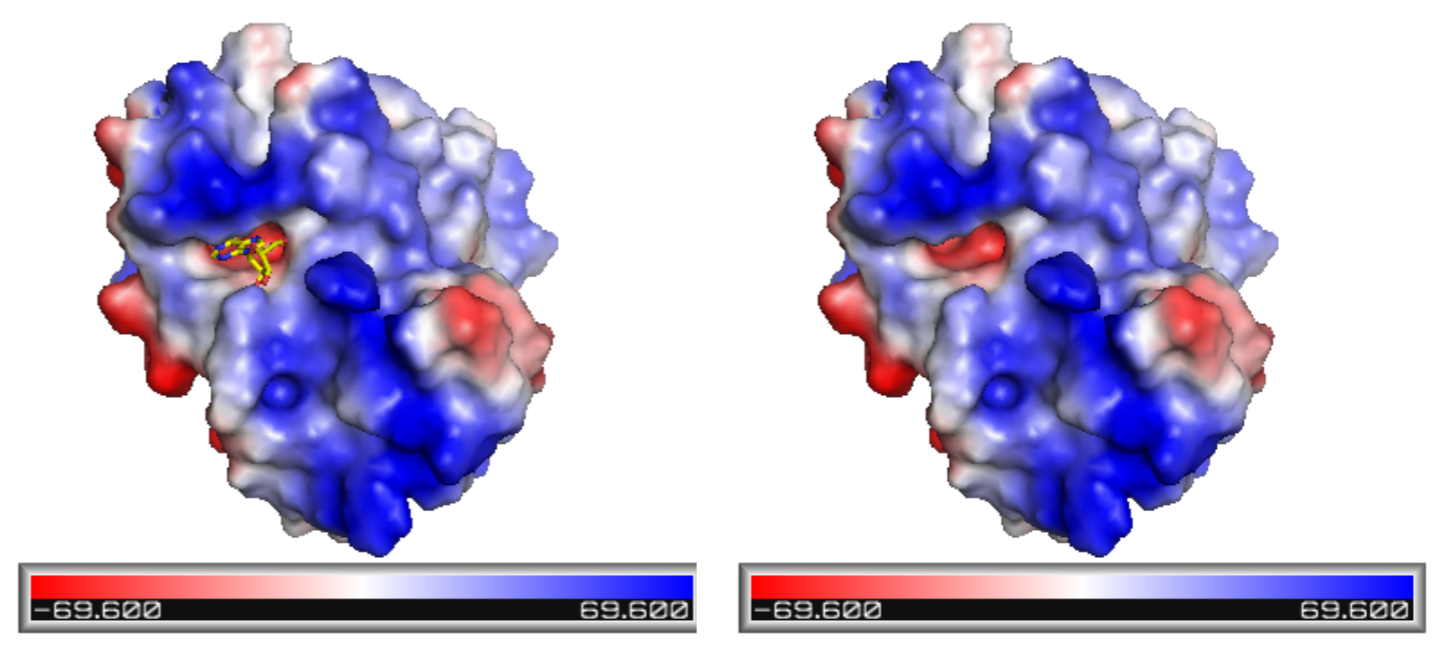


→ YTHDF2 with m6A (human)

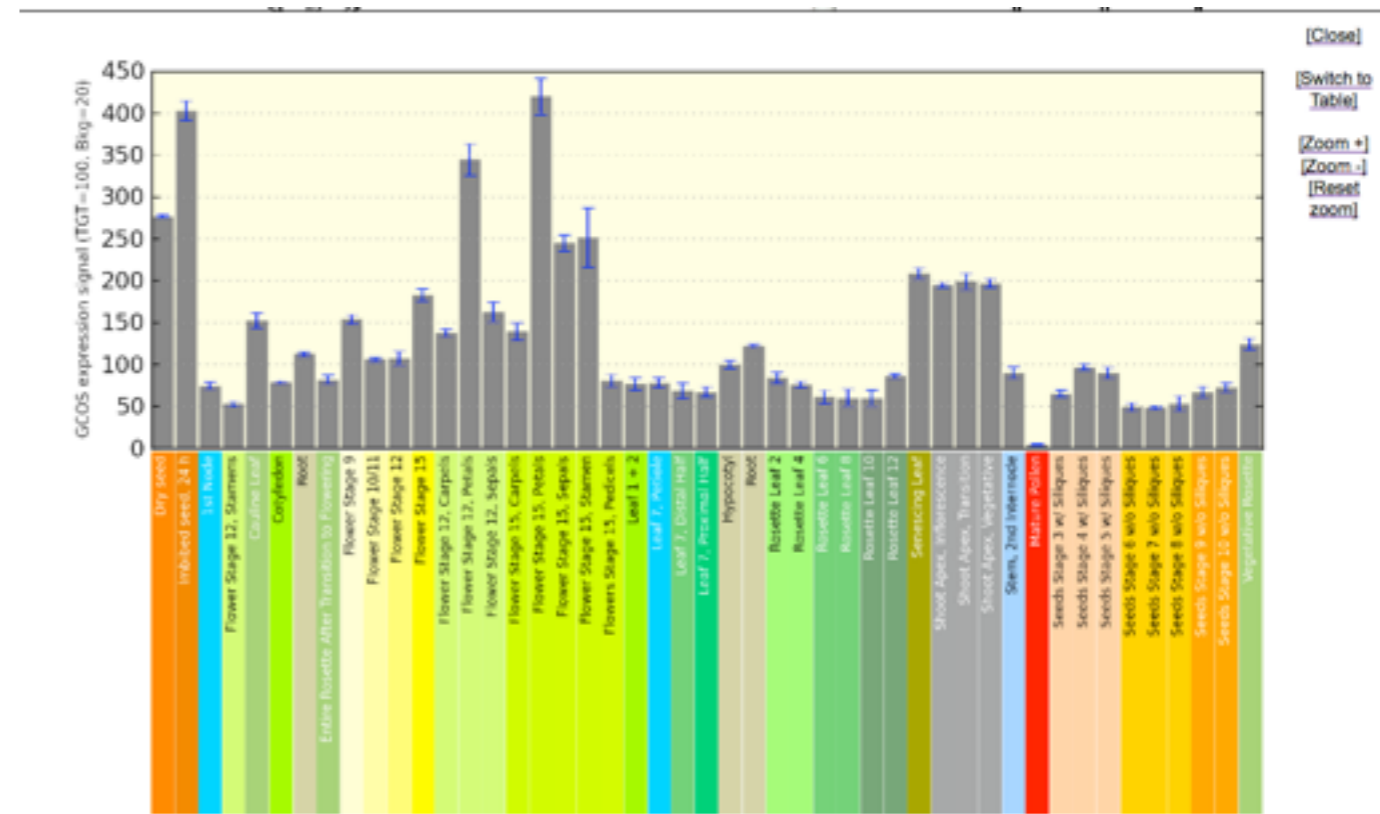
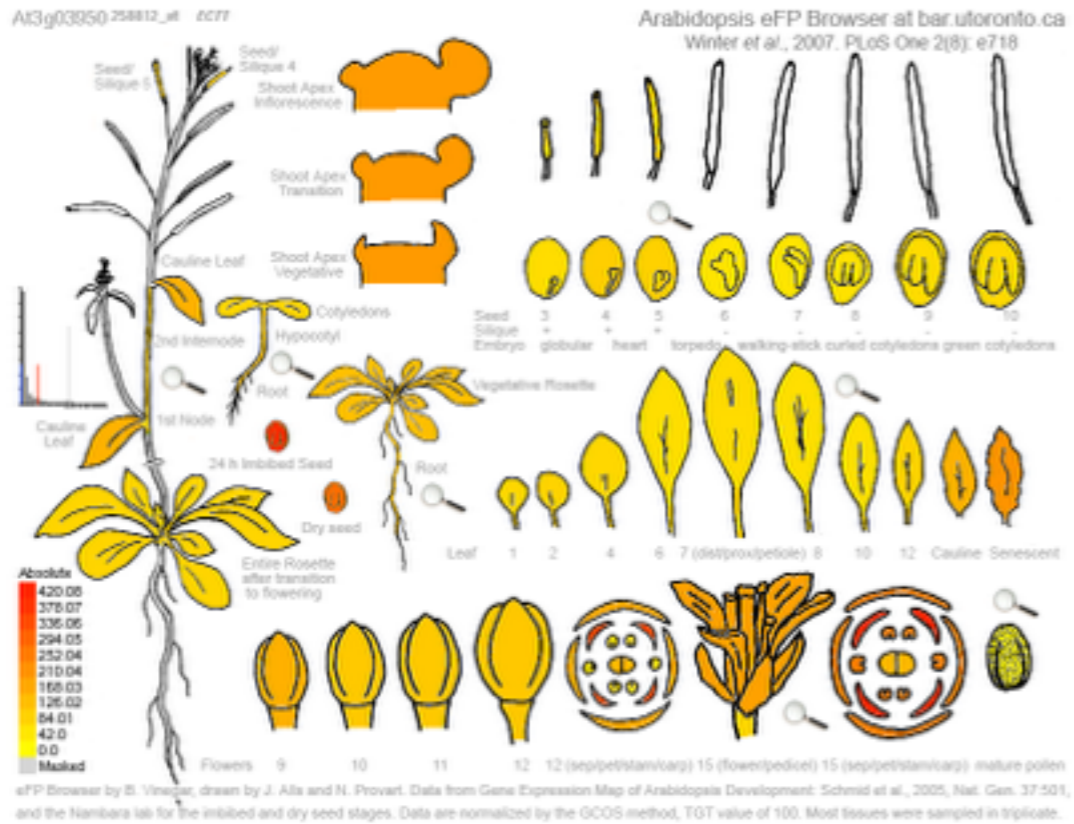
YTHDF2 magenta
AT3G03950 cyan



vaccum electrostatics
YTHDF2 AT3G03950



组织特异性表达



特异性表达

Baseline Expression

Display gradients
 Display levels
 Display variance

Gene	Bur-0	Can-0	Col-0	Col-1	Ecd-0	Hi-0	Kn-0	Ler-0	Mt-0	No-0	Oy-0	Po-0	Roch-4	Sf-2
ECT1	20	21	19	27	26	26	19	26	23	27	25	30	19	16

Differential Expression

Comparison	Log ₂ -fold change
'2 hour; excess light' vs '0 hour; low light' in 'wild type'	
'2 hour; excess light' vs '0 hour; low light' in 'tnr4 mutant'	
'excess light; none; 2 hour' vs 'low light; none; 0 hour'	
'2 hour; excess light' vs '0 hour; low light' in 'hsfa1dhsfa2hsfa3 mutant'	
'bzi1-1D;bri1-116 double mutant' vs 'wild type'	
'bzi1-1D;bri1-116 double mutant' vs 'bri1-116 null mutant'	
'37 degrees celsius' vs '20 degrees celsius'	
'low pH (pH 4.6)' vs 'standard pH (pH 5.7)' in 'root cortex'	
'low pH (pH 4.6)' vs 'standard pH (pH 5.7)' in 'root epidermis and lateral root cap'	
'Plasmodiophora brassicae' vs 'none' at '10 day'	
'oxf6:AtCPSF30' vs 'wild type'	
'0.5 hour; excess light' vs '0 hour; low light' in 'wild type'	
'fus3-3 mutant' vs 'wild type' in '12 day; Col-0'	
'primisulfuron' vs 'control'	
'sulfometuron methyl' vs 'control'	
'lec1-1 mutant' vs 'wild type' in '12 day; Ws'	
'low pH (pH 4.6)' vs 'standard pH (pH 5.7)' in 'root stele'	
'low pH (pH 4.6)' vs 'standard pH (pH 5.7)' in 'root endodermis and quiescent center'	
'imidazolinone' vs 'control'	
'excess light; none; 0.5 hour' vs 'low light; none; 0 hour'	
'50 micromolar; abscisic acid' vs 'control (ethanol vehicle)' in 'wild type genotype; leaf'	
'100 micromolar; 4-chloro-6-methyl-2-phenylpyrimidine' vs 'control' at '24 hour'	
'rrd2 mutant' vs 'wild type'	
'0.5 hour; excess light' vs '0 hour; low light' in 'tnr4 mutant'	
'triazolopyrimidine' vs 'control'	
'30% inhibition from optimum photosynthesis temperature' vs 'baseline growth temperature'	
'apl mutant' vs 'wild type'	
'lec1-1 mutant' vs 'wild type' in '8 day; Ws'	
'auxin' vs 'none' in '35 day'	
'low pH (pH 4.6)' vs 'standard pH (pH 5.7)' in 'root columella root cap'	



Your Input:

- ECT1 uncharacterized protein (428 aa)
(*Arabidopsis thaliana*)

Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
● AT2G39580	uncharacterized protein (1577 aa)				●					0.876
● CIPK1	CBL-interacting protein kinase 1; CIPK serine-threonine protein kinases interact with CBL prote [...] (444 aa)				●					0.787
● AT5G35910	exosome complex exonuclease RRP6 (870 aa)				●					0.674
● PAB3	poly(A) binding protein 3; Binds the poly(A) tail of mRNA (660 aa)				●					0.637
● AT2G32415	RRP6-like protein 3 (891 aa)				●					0.628
● AT1G54440	RRP6-like protein 1 (639 aa)				●					0.628
● CIPK17	CBL-interacting protein kinase 17; CIPK serine-threonine protein kinases interact with CBL pro [...] (432 aa)				●					0.588
● AT3G17740	uncharacterized protein (1149 aa)				●					0.587
● SOS2	SALT OVERLY SENSITIVE 2; Involved in the regulatory pathway for the control of intracellular Na [...] (446 aa)				●					0.578
● AT3G17712	uncharacterized protein (856 aa)				●					0.558

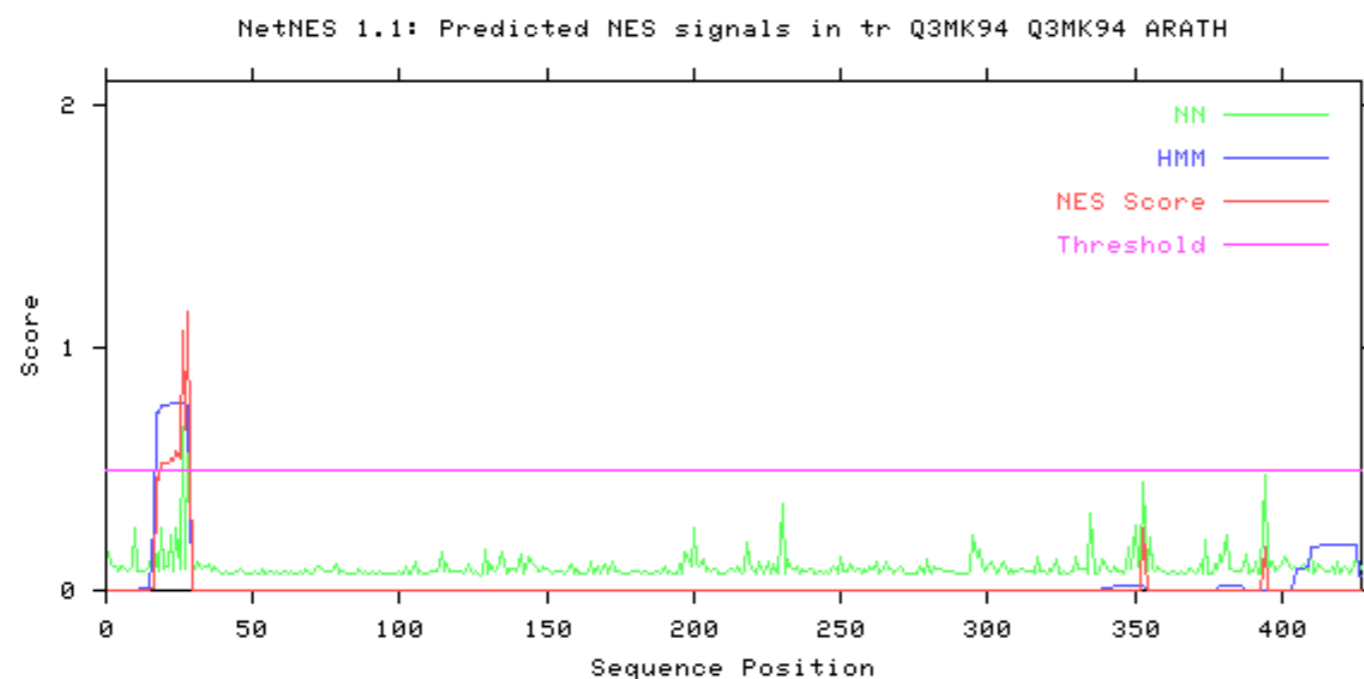
功能分析

定位

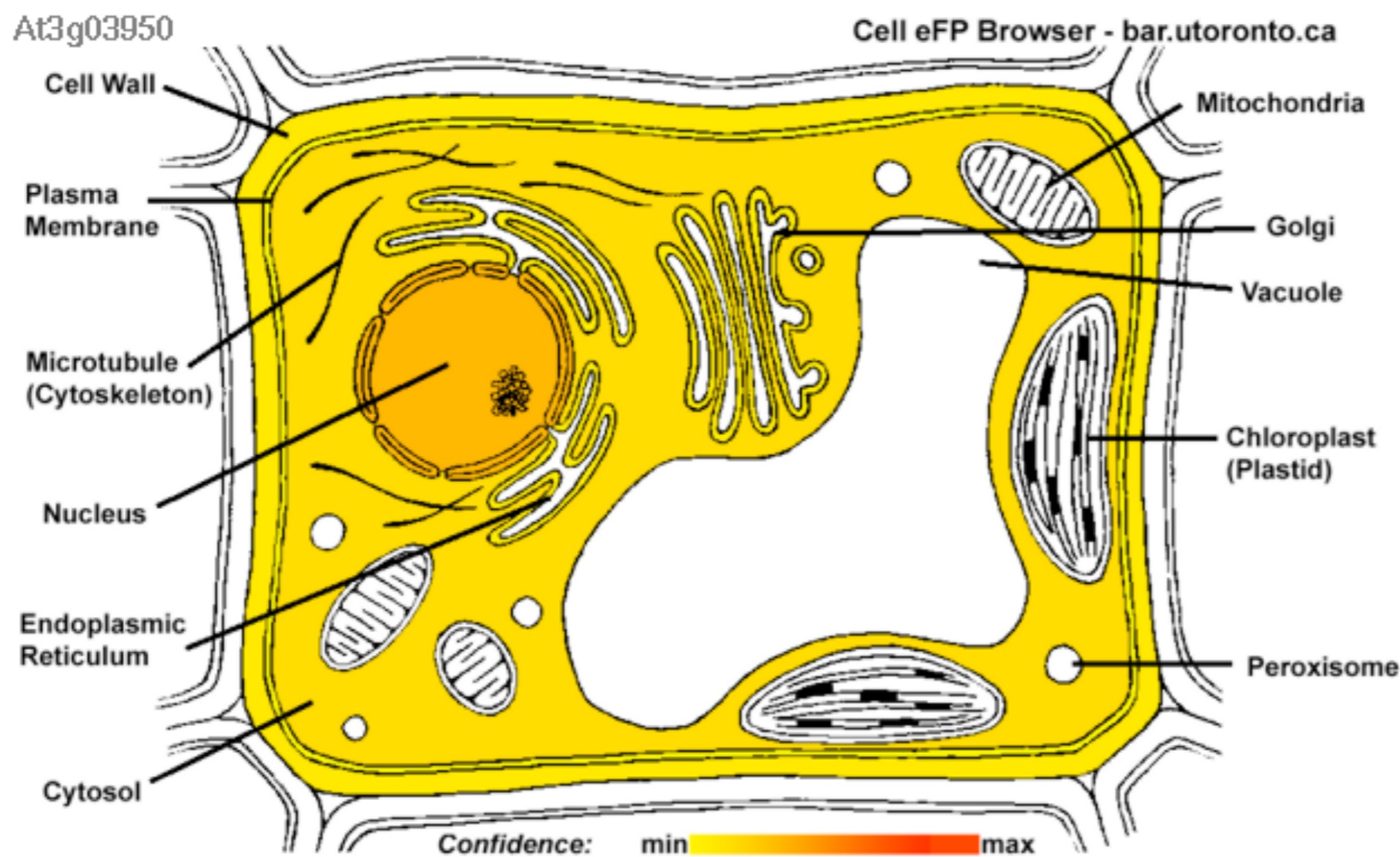
TargetP 1.1 Server

Name	Len	mTP	SP	other	Loc	RC
tr_Q3MK94_Q3MK94_ARA	428	0.069	0.136	0.797	_	2

NetNES 1.1 Server



tr_Q3MK94_Q3MK9-19-T	0.258	0.760	0.527	Yes
tr_Q3MK94_Q3MK9-20-A	0.094	0.760	0.528	Yes
tr_Q3MK94_Q3MK9-21-D	0.098	0.760	0.526	Yes
tr_Q3MK94_Q3MK9-22-L	0.225	0.772	0.545	Yes
tr_Q3MK94_Q3MK9-23-F	0.084	0.772	0.539	Yes
tr_Q3MK94_Q3MK9-24-Q	0.261	0.772	0.576	Yes
tr_Q3MK94_Q3MK9-25-D	0.091	0.772	0.542	Yes
tr_Q3MK94_Q3MK9-26-L	0.676	0.773	1.073	Yes
tr_Q3MK94_Q3MK9-27-S	0.086	0.762	0.660	Yes
tr_Q3MK94_Q3MK9-28-L	0.630	0.762	1.154	Yes



Drawn by T. Ampofo. Data from SUBA (Heazlewood et al, 2007).

Cell Details

Localisation Information

- cell plate
- cytoskeleton
- cytosol**
- endoplasmic reticulum
- extracellular
- golgi
- mitochondria
- nucleus**
- peroxisome
- plasma membrane**
- plastid
- unclear
- unknown
- vacuole

Data Source Options

Image Options

- Remove confidence gradient**

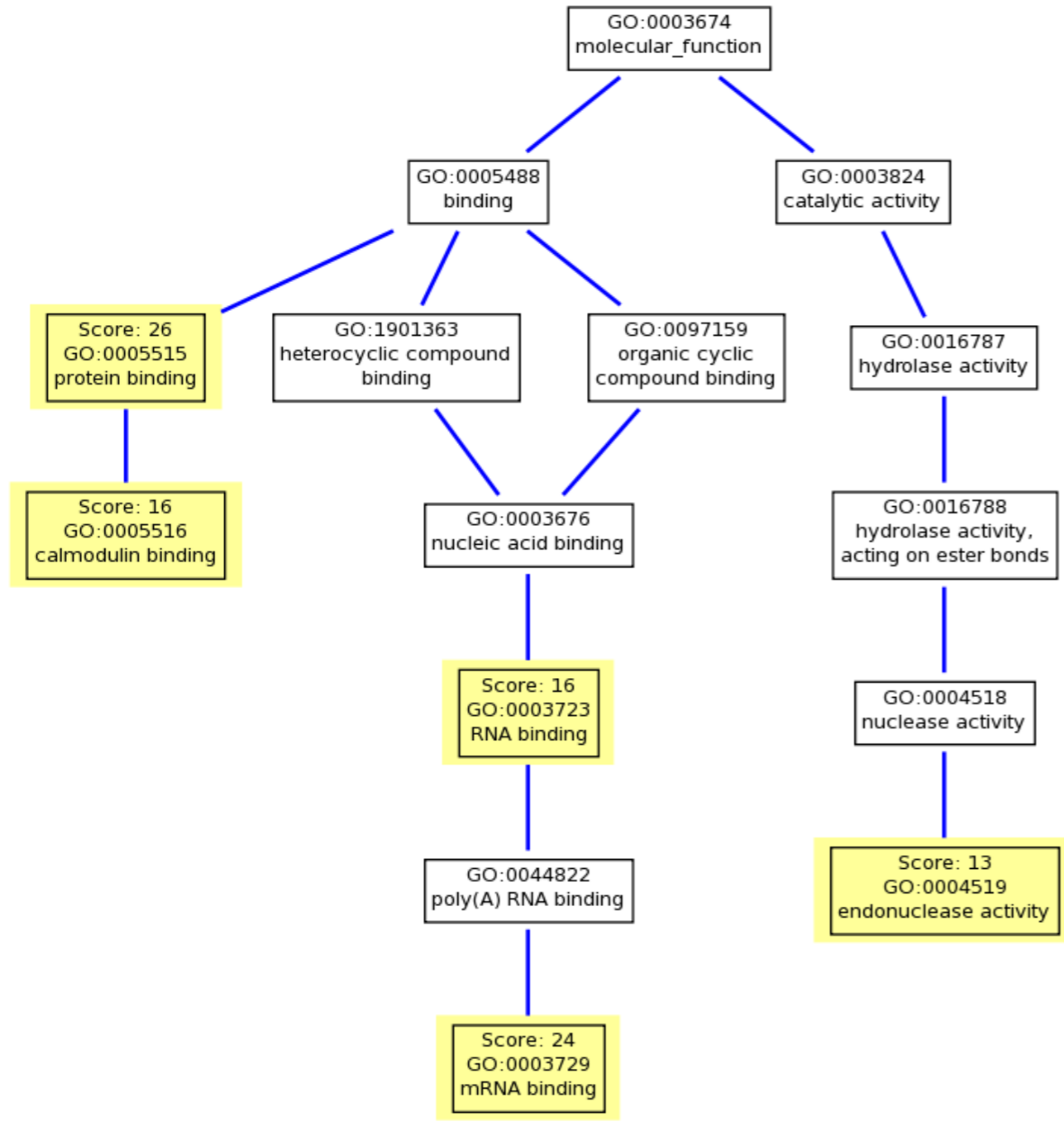
功能分析

Molecular Function Ontology				
#	GO ID	GO Term	Reliability (%)	<input checked="" type="checkbox"/>
1	GO:0005515	protein binding	26	<input checked="" type="checkbox"/>
2	GO:0003729	mRNA binding	24	<input checked="" type="checkbox"/>
3	GO:0003723	RNA binding	16	<input checked="" type="checkbox"/>
4	GO:0005516	calmodulin binding	16	<input checked="" type="checkbox"/>
5	GO:0004519	endonuclease activity	13	<input checked="" type="checkbox"/>

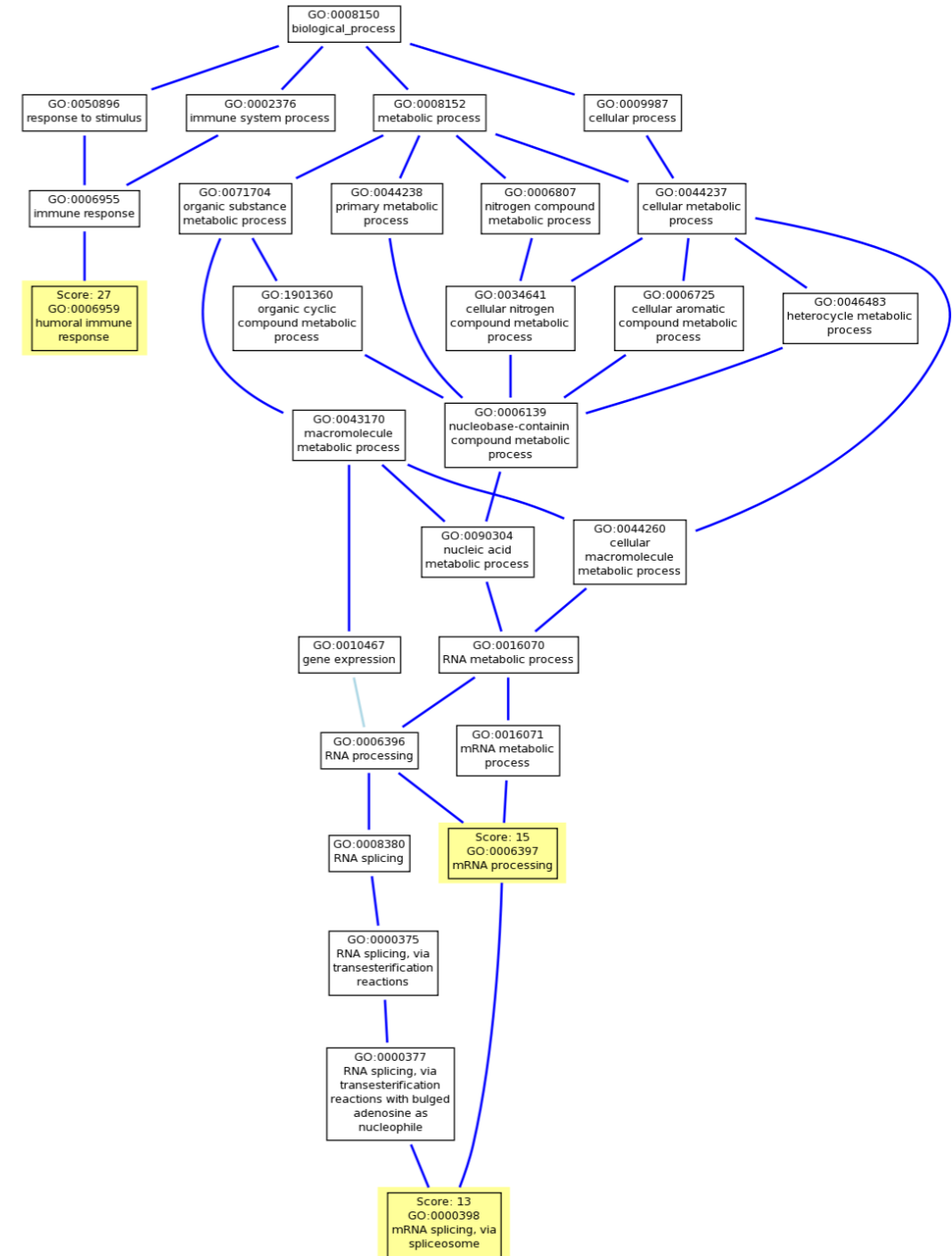
Biological Process Ontology				
#	GO ID	GO Term	Reliability (%)	<input checked="" type="checkbox"/>
1	GO:0006959	humoral immune response	27	<input checked="" type="checkbox"/>
2	GO:0006397	mRNA processing	15	<input checked="" type="checkbox"/>
3	GO:0000398	nuclear mRNA splicing, via spliceosome	13	<input checked="" type="checkbox"/>

功能分析

Molecular Function Ontology



Biological Process Ontology



总结

- 1 通过序列分析可知YTH结构域蛋白在拟南芥中含量丰富，可能具有重要的功能
- 2 通过结构分析可知ECT1蛋白C端的YTH结构域与人类YTHDF2非常相似，应该起到结合m6A的作用，而不同的N端结构域应该发挥不同的功能，且ECT1预测的溶剂可及性好，表达纯化可能相对容易
- 3 通过亚细胞定位，表达特异性，信号序列预测，相互作用蛋白等可以推测其功能

致谢

- 罗静初老师
- 张晨 贾桂芳
- 小组成员 魏连环 张健 张笑
- 全体同学

