



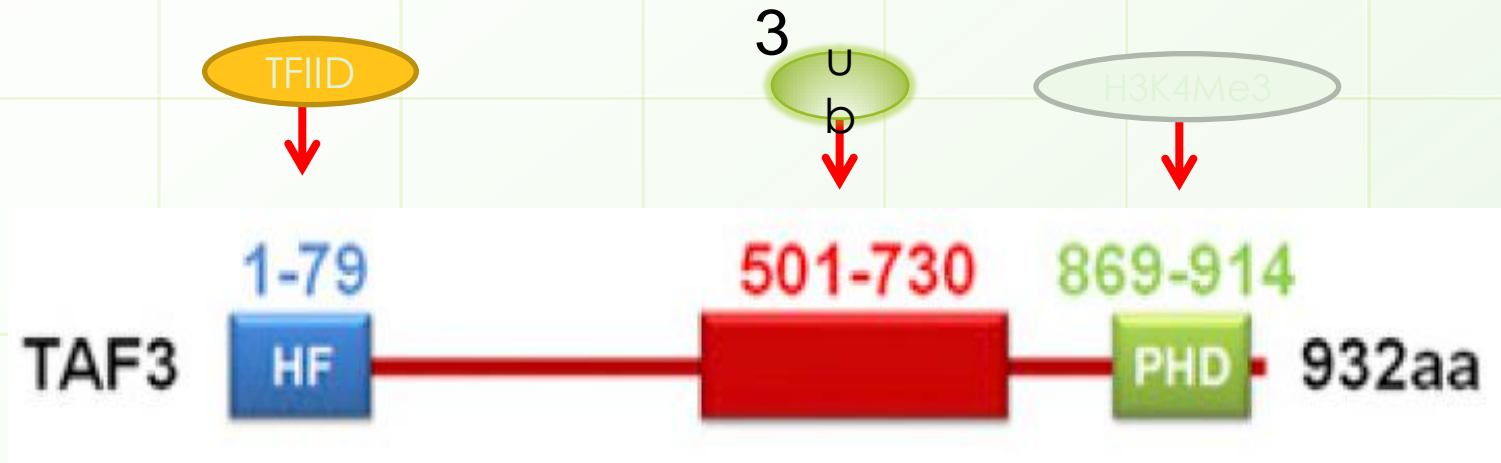
# 小鼠TAF3基因分析和 蛋白结构预测

**L/O/G/O** 2011-12-21

Reporter :

G05&G06

# Transcription initiation factor TFIID subunit



Vertebrate-specific region  
(1-79aa)

Histone Fold

(79aa)

PHD finger  
(869-914aa)

(501-730aa)

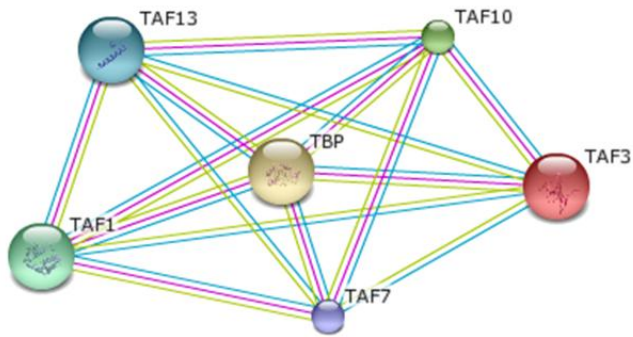
(1-79aa)

# Transcription initiation factor TFIID subunit 3

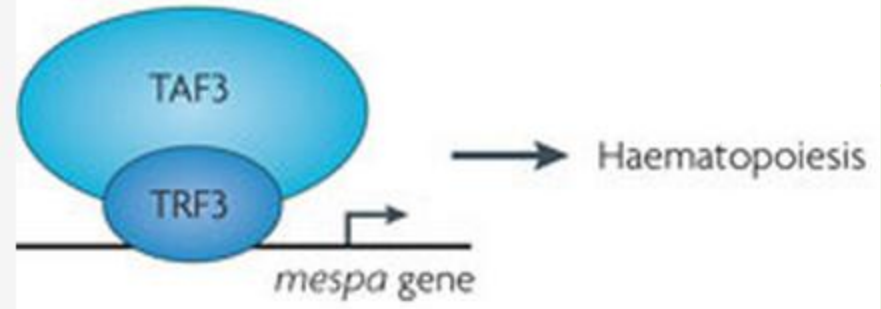
<http://biogps.org>

Symbol:	Taf3
Description:	TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated factor
Accessions:	<a href="#">209361</a> (NCBI Gene) <a href="#">ENSMUSG00000025782</a> (Ensembl) <a href="#">Q5HZG4</a> (UniProt) <a href="#">2388097</a> (MGI) <a href="#">35415</a> (HomoloGene)
Aliases:	140kDa, 4933439M23Rik, AW539625, TAF140, TAFII-140, TAFII140, mTAFII140
Genome Location:	<a href="#">chr2:9836179-9970223</a> (mm9)
Function:	<b>Molecular Function</b> p53 binding ( <a href="#">GO:0002039</a> ) protein binding ( <a href="#">GO:0005515</a> ) zinc ion binding ( <a href="#">GO:0008270</a> ) metal ion binding ( <a href="#">GO:0046872</a> ) <b>Biological Process</b> negative regulation of transcription from RNA polymerase II promoter ( <a href="#">GO:0000122</a> ) negative regulation of transcription from RNA polymerase II promoter ( <a href="#">GO:0000122</a> ) transcription, DNA-dependent ( <a href="#">GO:0006351</a> ) regulation of transcription, DNA-dependent ( <a href="#">GO:0006355</a> ) regulation of transcription from RNA polymerase II promoter ( <a href="#">GO:0006357</a> ) transcription from RNA polymerase II promoter ( <a href="#">GO:0006366</a> ) negative regulation of sequence-specific DNA binding transcription factor activity ( <a href="#">GO:0043433</a> ) negative regulation of sequence-specific DNA binding transcription factor activity ( <a href="#">GO:0043433</a> ) maintenance of protein location in nucleus ( <a href="#">GO:0051457</a> ) <b>Cellular Component</b> nucleus ( <a href="#">GO:0005634</a> ) nucleus ( <a href="#">GO:0005634</a> ) transcription factor TFIID complex ( <a href="#">GO:0005669</a> ) transcription factor TFIID complex ( <a href="#">GO:0005669</a> )

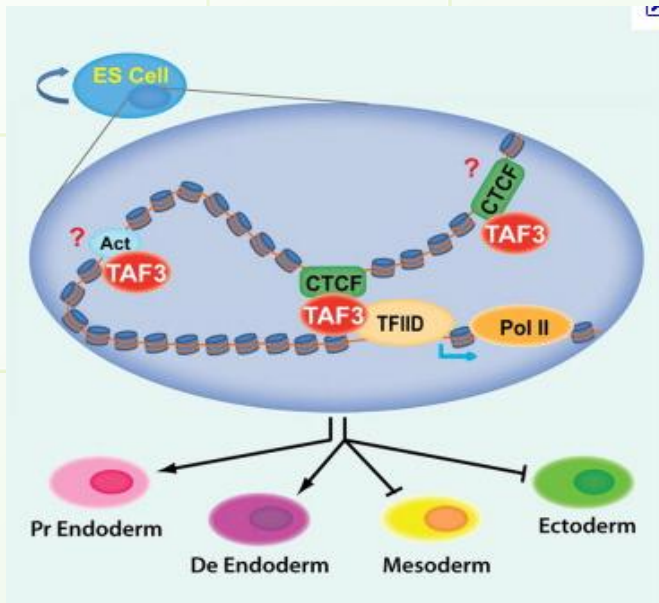
# Interaction & function



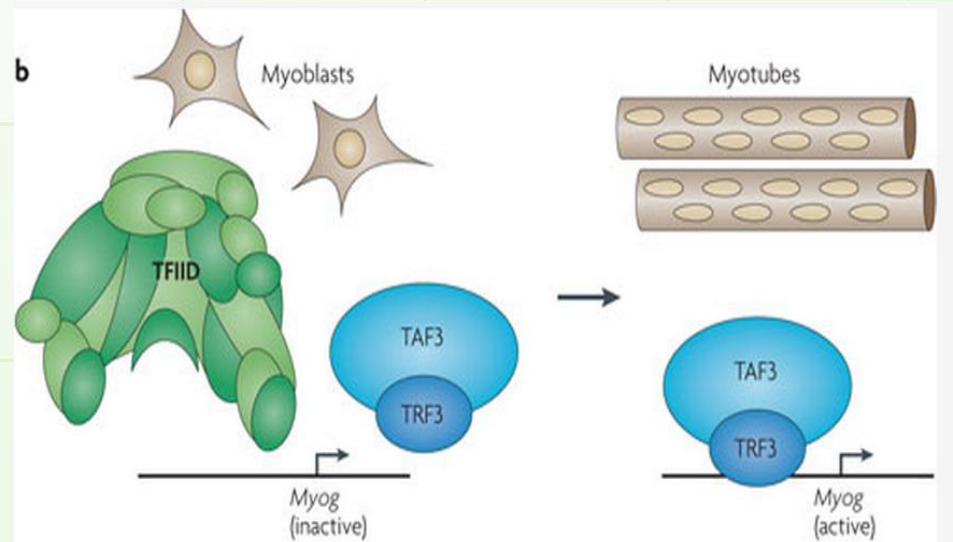
String 9.0



Hart, D. O., et al. *Nature* **450**, 1082–1085 (2007)



Yick W. Fong et al, *Cell*, 2011

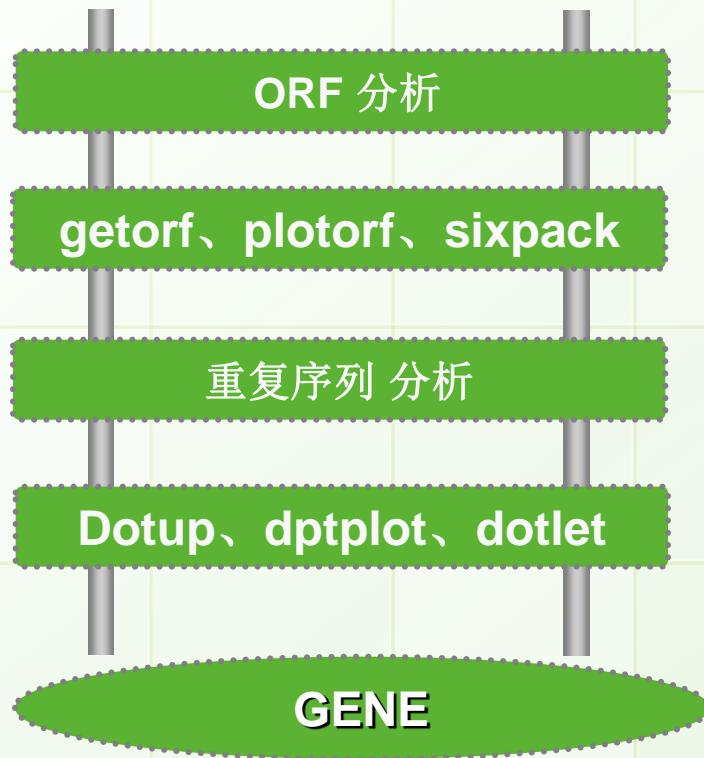


Nature Reviews | **Genetics**

Deato, M. D. et al., *R Genes Dev.* **21**, 2137–2149 (2007)

# TAF3 GENE & PROTEIN

## ANALYSIS AND PREDICTION



# ORF 分析 (getorf为例)

nucleotide size of ORF (500-10000) ORF最小设为500.

```
>NM_027748.3_1 [147 - 3005] Mus musculus TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated factor (Taf3), mRNA
GEAAGLESSGSEGLRWRPRSGMCEYSRSLLRVSVQAICQALGWDSVQLSACHLLTDVLQ
RYLQQLGRGCHRYSELYGRTDPILDDVGEAFQLMGVNLHELEDYIHNIEPVTFPHQIPSF
PVSKNNVLQFPQPGSKDAEERKDYIPDYLPPIVSSQEEEEEEQVPTDGGTSAEAMQVPLE
EDDEMEEEEEVINDENFLGKRPLDSPAEEEMPSMKRPRLLSTKGDSDLVVLEAREPLSSI
NPQKTFFVLSFVVRVQDRADLAPPSPQPPMLAFFAKSQLPIAKPLETKSFTPKTKKASSP
GQKTKSPKAALSPARLGSPIRSPKTIPEKKSPPGRSKSPKSPKSPKI VAHVPQTPVVRPET
PNRTPSAMVVEKTVKETIPVMKPTQTPPEVVKLNIEMQPKPVVTDKTIIDDSIDAVIARA
CAEREPDFFEFSSGSESEGDFTTSPKRISGSECATPKASTSSNNFTKSLATPLPLSSGTS
SSDNSWTMDASIDEVVRKAKLGAAPSNNPPTFFYISSPISPTPEPLHKGYEEKAKLPSS
VDVKKKLKKELKTKLKKKQKQDRERERERENKERSKEKDKMREREKEKEAGKELKYPWRE
LMKDESDPYKFKIKEFEDIDAAKVRLKDGIVRREKHKKDKKDRERSKREKDKRERER
LKEKNREDKIKAPPTQLVLPPEMALPLFSPSAVRVPAMLPAPSPMLPEKLFEEKEKPKKE
KERKKDKKKEKKEKEKEKEKEKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREK
VIRLTLRVGAGQDKIVISKVVPAPAEKAPAPSLNRPKTPPPAPVPIVVRVSPPLQPPLL
TQAAVCPALMPSAPALSIGSAKAPVRSVVTETVSTYVIRDEWGNQIWIICPGCNKPDG
SPMIGCDDDDWYHWPCVGI MAAPPEMQWFCPKCANKIKKDKKDKKDKKRAH
>NM_027748.3_2 [3501 - 2908] (REVERSE SENSE) Mus musculus TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated
factor (Taf3), mRNA
KSGEVGFSFIYSHHKNIPSAVTGAFFPSDYLPVEGVGLPTVATDAPFHQSRHLDITFLSNT
MTKSGGEEKGNREGRCVASLLILVTGNRQMEVPALGPALHWSFLQQAAGAQAHVQREH
PEGCHVLVLAFLGCAASPQGPGETGWQNPSPDPAQGSSPAHADVSAHGASSSCASCLSL
SCWHTSGRTTASLLEEQP
>NM_027748.3_3 [3257 - 2259] (REVERSE SENSE) Mus musculus TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated
factor (Taf3), mRNA
YLSLVTIGRWRCQPSGQSTGVFFSNMLRELRFMCFRGSIQRGAMFLSLLLWDALQVHRAR
GRQDGRTPSPAPTRRRAAVLPTLMSVRTVLPILLVLLVFLYLVTGTLRAEPLHLFWRSSHDSH
TGPVVPIVTVVTANHRAPVIRLVAPWTDPDLIAPLITDHISADRLSHDAPHGGFGTTPYSR
EGGSWGHQGWAMCSLGEWVWLQGSRAHSHRNRDHWGRWGFSGVQRRSGLSLRGWYHLAD
NNLVLASANSQSQPGDNWTWGRDDRLHFYLLMFVLLLLLLLSLLSPLSLLPLPLFLLLFF
LLLLLLLLLFLVLPFFFLGLLFFLKQFFWEHW
>NM_027748.3_4 [2776 - 1976] (REVERSE SENSE) Mus musculus TAF3 RNA polymerase II, TATA box binding protein (TBP)-associated
factor (Taf3), mRNA
VLTVSVTLLTGALALPIPERAGAGDGIRAGHTAAVRSGGCRGVGLTRTGIGTGAGGGV
LGLFREGAGLASGAGTLLITILSWPAPTRRVNLGITGLGAGMTGSTFIFSCLCFSFSPS
LPSLSLLSFLSLSLFFSFSFSSFFSFFSFLSFFLSFSLGFSFSSNSFSGSIGEKAGSMA
GTLTAEGLNKGSAISFGGNTSCVGGAFILSSLFFSLSLSLRSLSRLLLSRSFFLSLFC
SRSLLTIPSFNRTFAASMSNSLILNL
```

# ORF 分析 (getorf为例)

		gi 262231747 ref NM 027748.3																								
1	AGCGGGAGGC GGAGCGAGTC CAAAATGGCG GCTCTCAGGC TGGCGCGCTC CGTGCTGCTG AGGCTTTGAG GTGGTCGCTC CGGGTCGGAG GGGGGACGAT TCGCCCTCCG CCTCGCTCAG GTTTTACCGC CGAGAGTCCG ACCGCGCGAG GCACGACGAC TCCGAAACTC CACCAGCGAG GCCCAGCCTC CCCCTGCTA																									
+3		Met	Asn	Arg	Gly	Leu	Cys	***	Gly	Glu	Ala	Ala	Gly	Leu	Glu	Ser	Ser	Gly	Ser	Glu	Gly	Leu	Arg	Trp	Arg	Pro
101	TTCCCCGCCG CGGGGCCCCC AGAGAATGAA TCGGGGGCTC TGCTGAGGCG AGGCGGCAGG GCTGGAGAGC AGTGGCAGCG AAGGGCTGCG GTGGCGTCCA AAGGGGCGGC GCCCCGGGGG TCTCTTACTT AGCCCCGAG ACGACTCCGC TCCGCCGTCG CGACCTCTCG TCACCGTCGC TTCCCGACGC CACCGCAGGT																									
+3	Arg Ser Gly																									
201	CGCAGCGGGA TGTGCGAGAG TTACTCTAGG TCGTTGTTGA GGGTCTCGGT GCGCGAGATC TGCCAGGCGC TGGGCTGGGA CTCGGTGCAG CTCAGCGCCT GCGTCGCCCT ACACGCTCTC AATGAGATCC AGCAACAAC TCCAGAGCCA CCGCGTCTAG ACGGTCCGCG ACCCGACCCT GAGCCACGTC GAGTCGCGGA																									

结果分析:

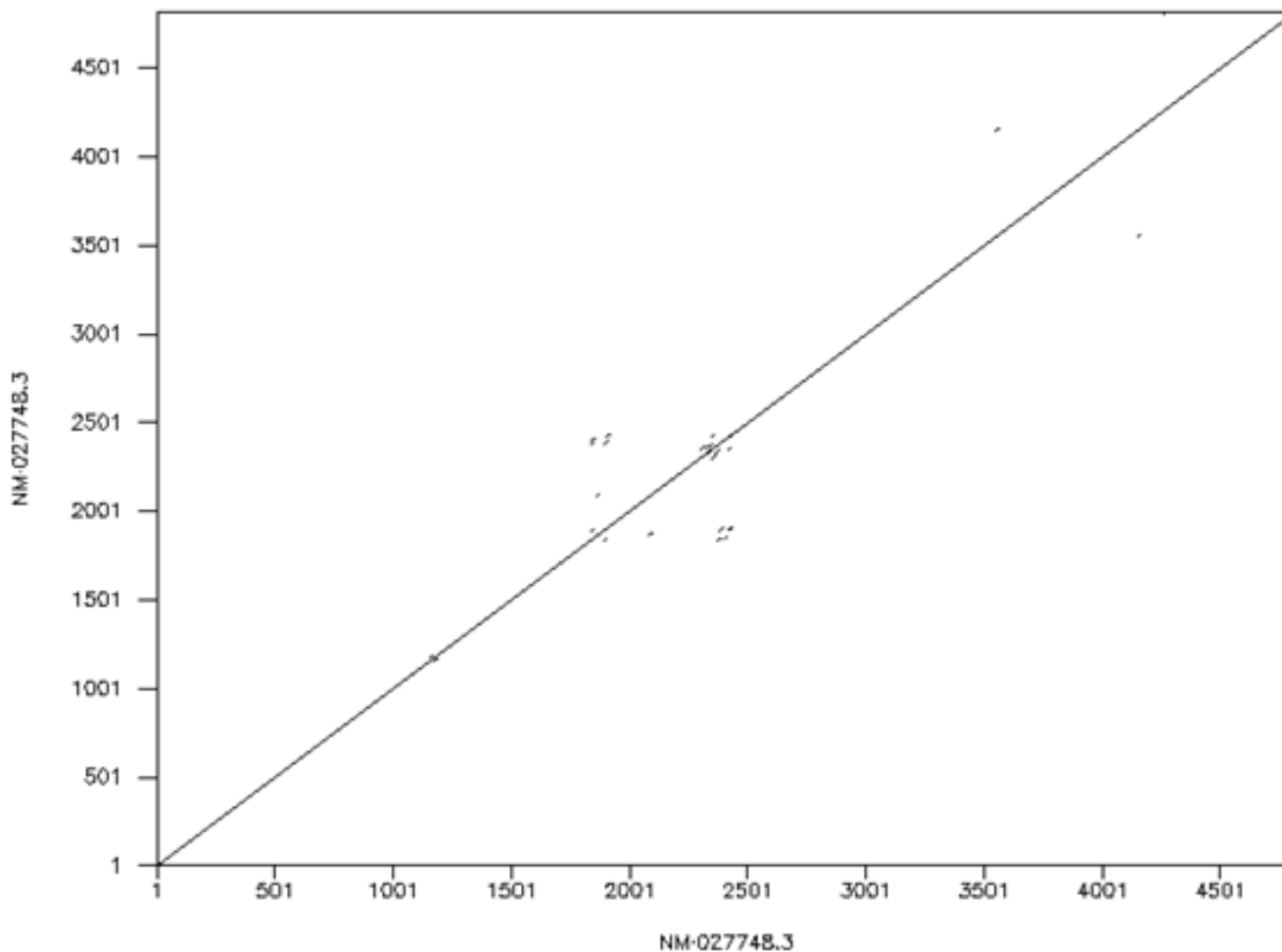
可见第一个预测结果包含TAF3的CDS区,但是由于其前方第126-128位为ATG,且恰好翻译到CDS的第一个氨基酸(209-211)。故二者连接在一起,被预测出来,可见这种隐性预测结果还是不能避免,这是这种程序共有的缺陷,需要生物学的方法进一步鉴定。

# 重复序列分析（dottup为例）

Word size: 12

Dottup: fasta::493175:NM-027748.3 vs fasta::493175:NM-02...

Wed 21 Dec 2011 19:43:05



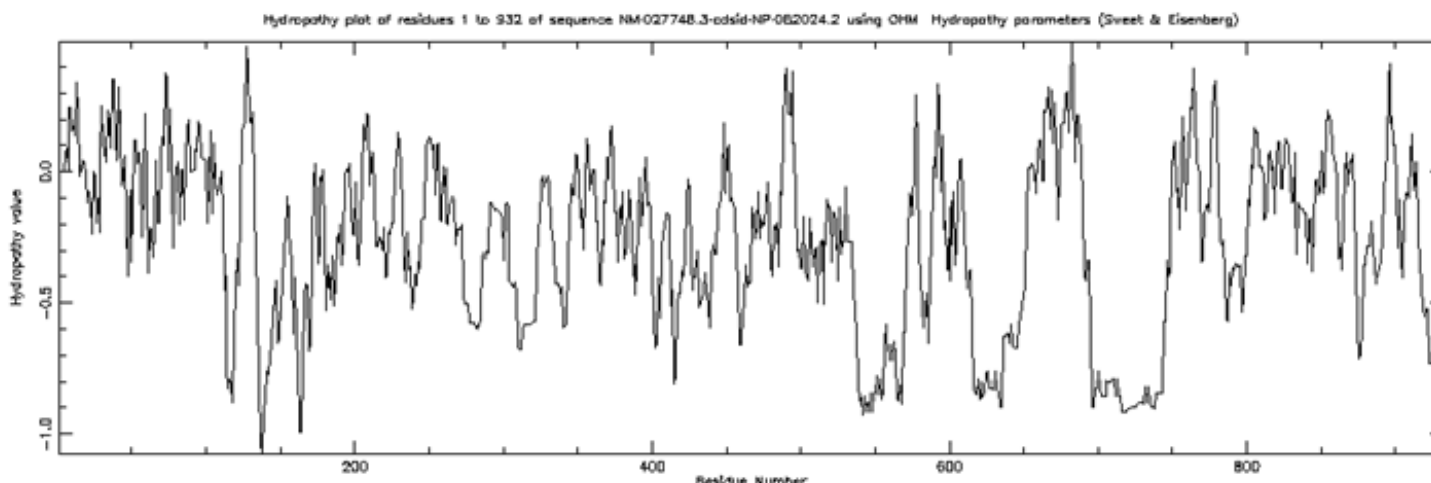
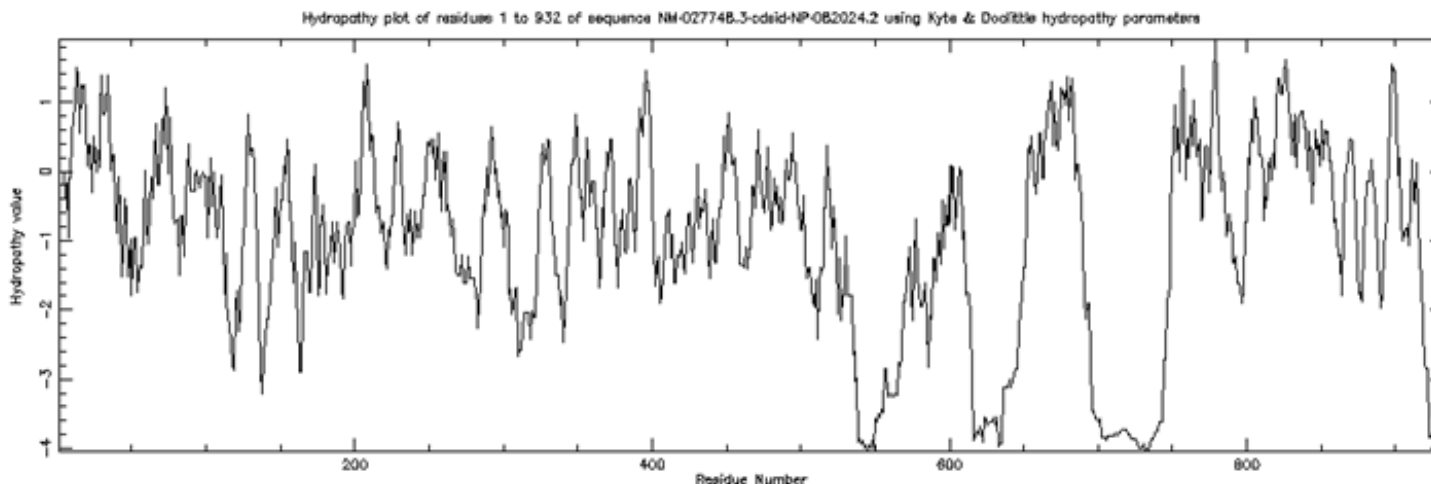
结果分析：  
可见TAF3-MUS  
mRNA中没有长的  
重复序列，

只是在2001-  
2501区有较多的  
12个碱基重复。



氨基酸特性分析—疏水性

Pepinfo (V6.0.1)



Hydropathy plot of residues 1 to 932 of sequence NM-027748.3-oidid-NP-082024.2 using Consensus parameters (Eisenberg et al)

疏水性分析:  
在550-750  
这段区域出  
现了三处疏  
水区。对应  
的是脊椎动  
物特有区域。  
富含赖氨酸  
Lys。

Compositional bias 508 - 749 242 Lys-rich



## 氨基酸特性分析—构成

pepstats(v6.0.1)

PEPSTATS of NM\_027748.3\_cdsid\_NP\_082024.2 from 1 to 932

Molecular weight = 105114.55                      Residues = 932  
 Average Residue Weight = 112.784                Charge = 37.0  
 Isoelectric Point = 9.8146  
 A280 Molar Extinction Coefficient = 62160  
 A280 Extinction Coefficient 1mg/ml = 0.59  
 Improbability of expression in inclusion bodies = 0.909

Residue	Number	Mole%	DayhoffStat
A = Ala	55	5.901	0.686
B = Asx	0	0.000	0.000
C = Cys	14	1.502	0.518
D = Asp	50	5.365	0.975
E = Glu	97	10.408	1.735
F = Phe	18	1.931	0.536
G = Gly	31	3.326	0.396
H = His	14	1.502	0.751
I = Ile	39	4.185	0.930
J = ---	0	0.000	0.000
K = Lys	119	12.768	1.935
L = Leu	61	6.545	0.884
M = Met	21	2.253	1.325
N = Asn	19	2.039	0.474
O = ---	0	0.000	0.000
P = Pro	112	12.017	2.311
Q = Gln	28	3.004	0.770
R = Arg	58	6.223	1.270
S = Ser	77	8.262	1.180
T = Thr	41	4.399	0.721
U = ---	0	0.000	0.000
V = Val	57	6.116	0.927
W = Trp	8	0.858	0.660
X = Xaa	0	0.000	0.000
Y = Tyr	13	1.395	0.410
Z = Glx	0	0.000	0.000

可以看出含量较多的氨基酸有：Lys 赖氨酸119个；Pro 脯氨酸112个；Glu 谷氨酸97个。赖氨酸属于碱性氨基酸；谷氨酸属于酸性氨基酸；脯氨酸有利于形成转角。另外等电点为9.8，是碱性蛋白，有利于结合DNA。

# 重复序列分析

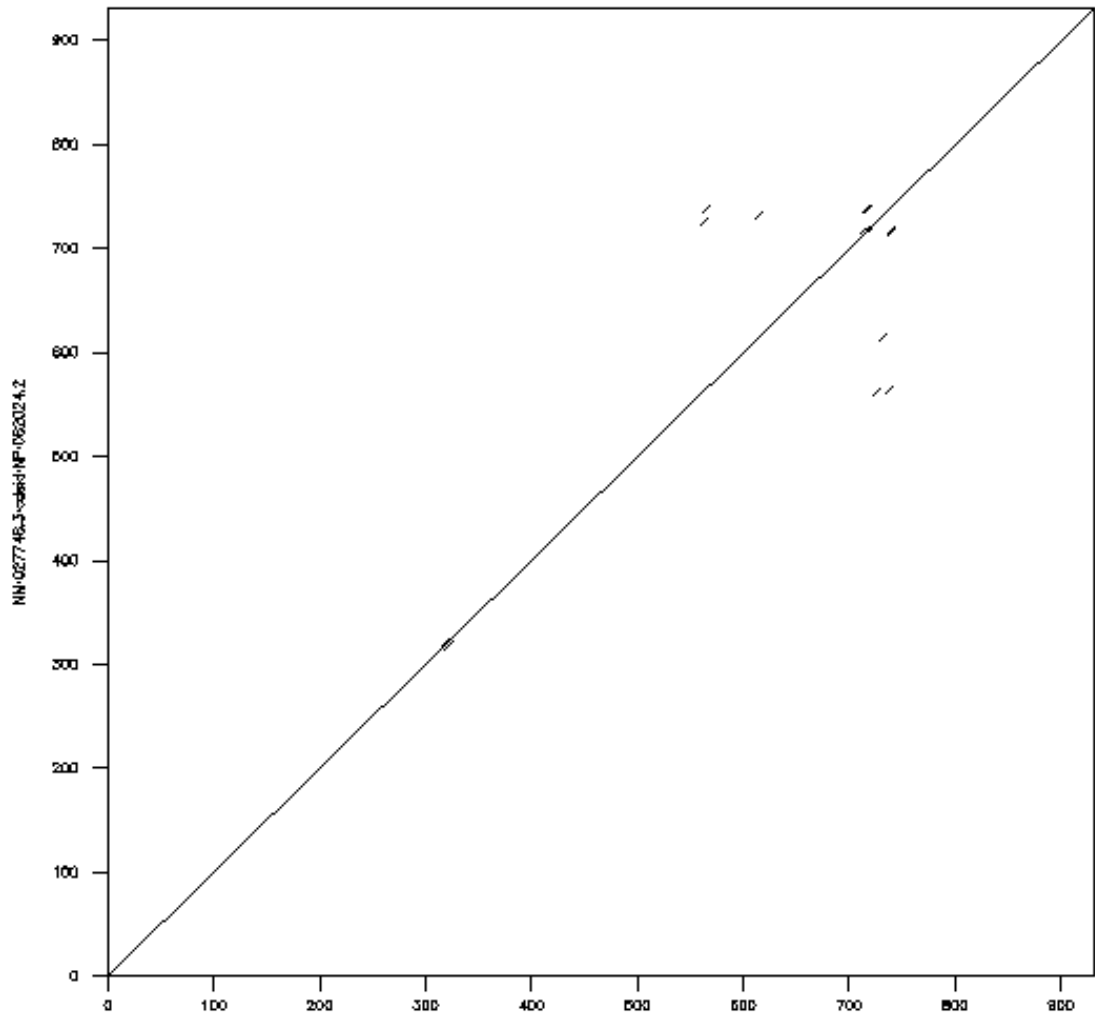
## Polydot V 6.0.1

### PROTEIN

WORD SIZE : 6

Poly dotplot of 493176

Thu 22 Dec 2011 03:30:08



No.	Length	Lines	Points	Sequence
1	932	15	1020	NW-027748.3-cds1d-NP-082024.2

NW-027748.3-cds1d-NP-082024.2

## 保守区域预测分析

取以下不同物种的TAF3序列做motif预测。（PSI BLAST 4次）

- 1. sp|Q5HZG4|TAF3 MOUSE Transcription initiation factor TFIID subunit 3 OS Mus musculus GN Taf3 PE 1 SV 2
- 2. tr|A2ASY1|A2ASY1 MOUSE TAF3 RNA polymerase II TATA box binding protein (TBP)-associated factor OS Mus musculus GN Taf3 PE 2 SV 1
- 3. tr|A2ASY0|A2ASY0 MOUSE TAF3 RNA polymerase II TATA box binding protein (TBP)-associated factor OS Mus musculus GN Taf3 PE 4 SV 1
- 4. tr|D3ZPB7|D3ZPB7 RAT RCG55818 OS Rattus norvegicus GN LOC100380100 PE 4 SV 1
- 5. tr|G3GR80|G3GR80 CRIGR Transcription initiation factor TFIID subunit 3 OS Cricetulus griseus GN I79 000003 PE 4 SV 1
- 6. tr|Q05BL5|Q05BL5 MOUSE Taf3 protein (Fragment) OS Mus musculus GN Taf3 PE 2 SV 1
- 7. tr|Q8BKP6|Q8BKP6 MOUSE Putative uncharacterized protein (Fragment) OS Mus musculus GN Taf3 PE 2 SV 1
- 8. gj|74747393|sp|Q5VWG9.1|TAF3 HUMAN RecName: Full Transcription initiation factor TFIID subunit 3 AltName: Full 140 kDa TATA box-binding protein-associated factor AltName: Full TBP-associated factor 30
- 9. gj|82075449|sp|Q5F489.1|TAF3 CHICK RecName: Full Transcription initiation factor TFIID subunit 3 AltName: Full TBP-associated factor 3
- 10. gj|82178917|sp|Q5EAW9.1|TAF3 XENLA RecName: Full Transcription initiation factor TFIID subunit 3 AltName: Full TBP-associated factor 3

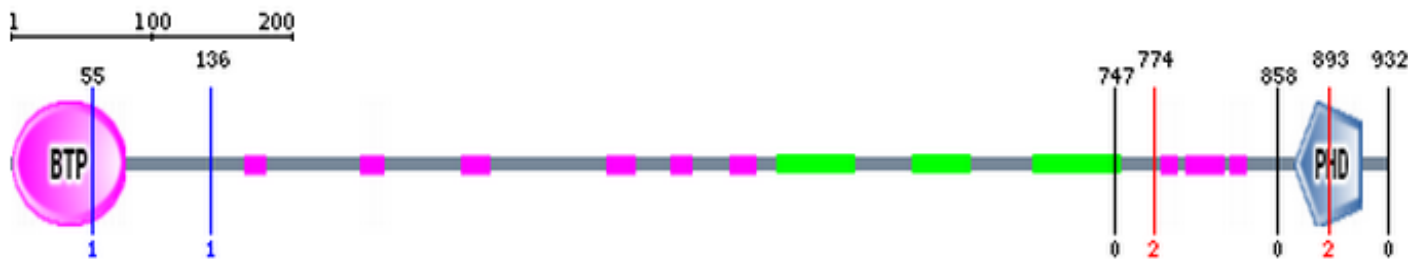
# 保守区域预测分析—SMART



[SETUP](#)
[FAQ](#)
[ABOUT](#)
[GLOSSARY](#)
[WHAT'S NEW](#)
[FEEDBACK](#)

## Domains within *Mus musculus* protein TAF3\_MOUSE (Q5HZG4)

Transcription initiation factor TFIIID subunit 3



Confidently predicted domains, repeats, motifs and features:

Name	Begin	End	E-value
BTP	3	79	1.94e-34
low complexity	159	173	-
low complexity	237	253	-
low complexity	306	325	-
low complexity	404	423	-
low complexity	447	461	-
low complexity	487	505	-
coiled coil	519	572	-
coiled coil	611	651	-
coiled coil	692	751	-
low complexity	779	790	-
low complexity	795	821	-
low complexity	826	837	-
PHD	869	915	4.77e-11

# 保守区域预测分析—MEME

## Motifs ?

The following motifs were supplied to MAST from "/.meme.html" last modified on Wed Dec 21 08:55:48 2011.

Motif	Width	Best possible match	Similarity		
			1	2	3
1	41	VSTYVIRDEWGNQIWIICFGCNKPDGSPMIGCDDCDDWYHW	-	0.23	0.15
2	50	DKTIDDSIDAVIARACAREPDPPEFSSGSESGDTFTSPKRISGSECTT	0.23	-	0.15
3	50	MQVPLEEDDEMEEEVINDENFLGKRPLDSPVEBEMPSMKRPRLLSTKGD	0.15	0.15	-

### Motif Overview

#### Motif 1

- 2.0e-268
- 10 sites



#### Motif 2

- 5.9e-326
- 10 sites



#### Motif 3

- 4.4e-271
- 10 sites

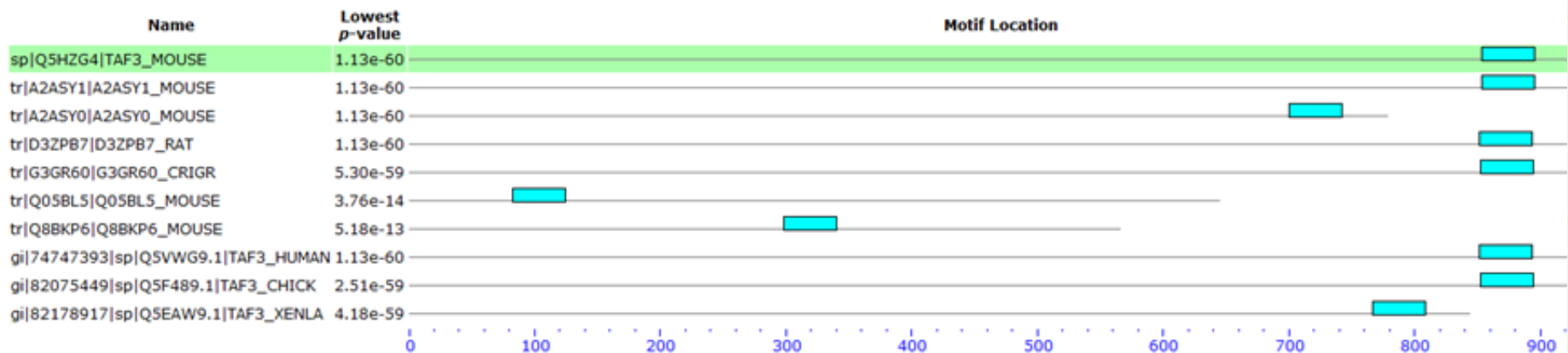


# 保守区域预测分析—MEME

Motif1对应的位点：853-894，对应PHD domain，结合H3K4me3.

Click on any row to highlight sequence in all motifs.

Name	Start	p-value	Sites
gi 74747393 sp Q5VWG9.1 TAF3_HUMAN	851	1.13e-60	APVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDDCDDWYHW PCVGIMTAPP
tr D3ZPB7 D3ZPB7_RAT	851	1.13e-60	APVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDDCDDWYHW PCVGIMAAPP
tr A2ASY0 A2ASY0_MOUSE	700	1.13e-60	APVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDDCDDWYHW PCVGIMAAPP
tr A2ASY1 A2ASY1_MOUSE	853	1.13e-60	APVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDDCDDWYHW PCVGIMAAPP
sp Q5HZG4 TAF3_MOUSE	853	1.13e-60	APVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDDCDDWYHW PCVGIMAAPP
gi 82075449 sp Q5F489.1 TAF3_CHICK	852	2.51e-59	APVRSVVTET VSTYVIRDEWGNQIWFCPGCNKDDGS MIGCDDCDDWYHW PCVGITAAPP
gi 82178917 sp Q5EAW9.1 TAF3_XENLA	767	4.18e-59	TPVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDQCDDWYHW PCVGINAAPP
tr G3GR60 G3GR60_CRIGR	852	5.30e-59	APVRSVVTET VSTYVIRDEWGNQIWICPGCNKDDGS MIGCDGCDDWYHW PCVGLMAAPP
tr Q05BL5 Q05BL5_MOUSE	82	3.76e-14	MGVNLHELED YIHNIETVTFPHQIPSPVSKNNVLQFQPGSKDAERKDY IPDYLPPIVS
tr Q8BKP6 Q8BKP6_MOUSE	298	5.18e-13	GSECATPKAS TSSNNFTKSLATLPLSSGTSSSDNSWTMDASIDEVVVKAK LGAPSNMPPT

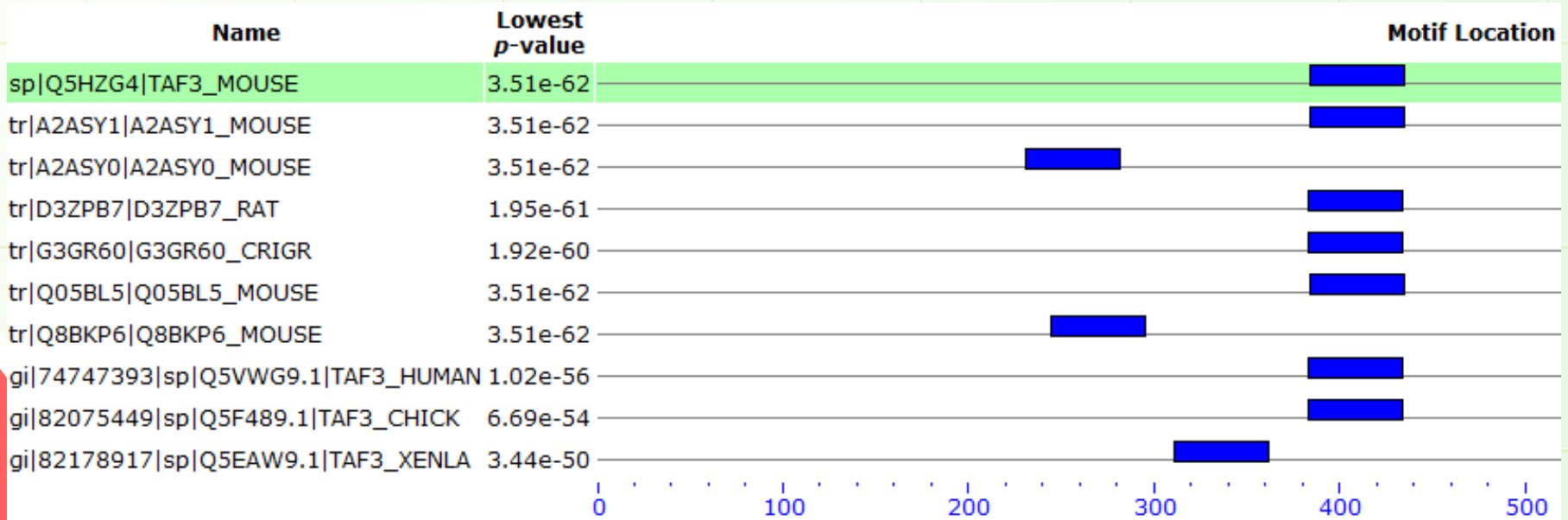


# 保守区域预测分析—MEME

Motif2对应的位点: 384-434 (未知区域)

Click on any row to highlight sequence in all motifs.

Name	Start	p-value	Sites
tr Q8BKP6 Q8BKP6_MOUSE	244	3.51e-62	EMQPKKPVVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGDFTTSPKRI SGSECAT PKASTSSNNF
tr Q05BL5 Q05BL5_MOUSE	384	3.51e-62	EMQPKKPVVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGDFTTSPKRI SGSECAT PKASTSSNNF
tr A2ASY0 A2ASY0_MOUSE	231	3.51e-62	EMQPKKPVVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGDFTTSPKRI SGSECAT PKASTSSNNF
tr A2ASY1 A2ASY1_MOUSE	384	3.51e-62	EMQPKKPVVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGDFTTSPKRI SGSECAT PKASTSSNNF
sp Q5HZG4 TAF3_MOUSE	384	3.51e-62	EMQPKKPVVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGDFTTSPKRI SGSECAT PKASTSSNNF
tr D3ZPB7 D3ZPB7_RAT	383	1.95e-61	EMQLKKGAVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGDFTTSPKRI SVSECTT PKASTSSNNF
tr G3GR60 G3GR60_CRIGR	383	1.92e-60	EMQPKKAVVT DKTIDDSIDAVIARACAEREPPDFEFSSGSESEVDFTTSPKRI SGSECTT PKASTSSNNF
gi 74747393 sp Q5VWG9.1 TAF3_HUMAN	383	1.02e-56	ENQPKKAVVA DKTIEASIDAVIARACAEREPPDFEFSSGSESEGDIFTSPKRI SGPECTT PKASTSANNF
gi 82075449 sp Q5F489.1 TAF3_CHICK	383	6.69e-54	ENQTKKLPVV DKTIDDSIDAVIARACAEREPPDFEFSSGSESEGEIFTSPKRLSVSETTA TTPKPSVSTN
gi 82178917 sp Q5EAW9.1 TAF3_XENLA	311	3.44e-50	EIPSKKPSIA DNTIEDSIDAVIARACAEREPPDFEFSSGSESEGEVFTSPKRLNISLTT PKVASAGINP



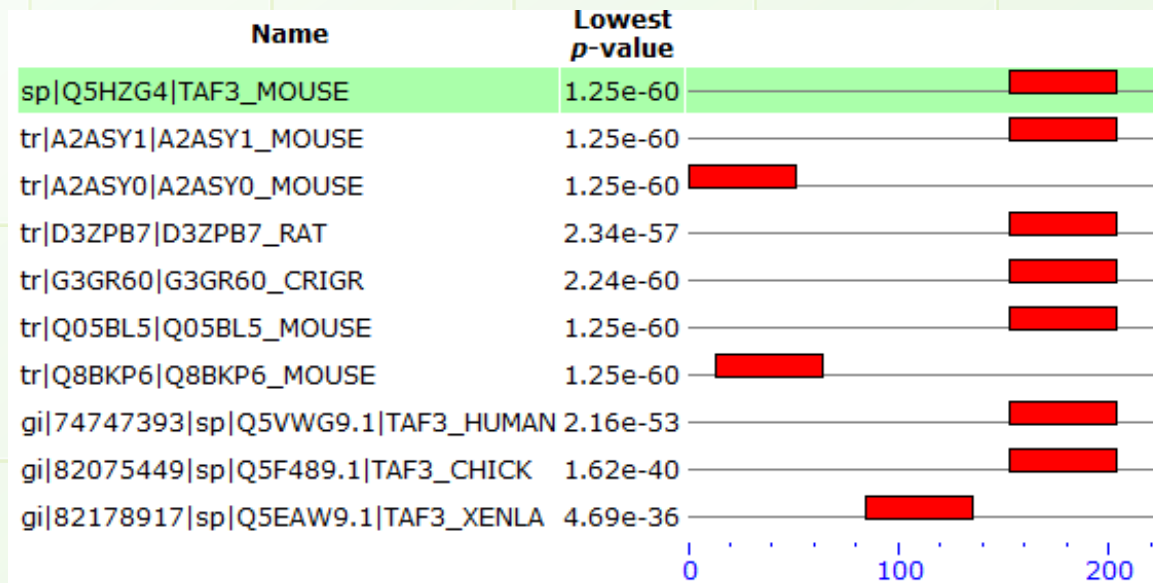


# 保守区域预测分析—MEME

## Motif3对应的位点: 153-203 (未知区域)

Click on any row to highlight sequence in all motifs.

Name	Start	p-value	Sites
tr Q8BKP6 Q8BKP6_MOUSE	13	1.25e-60	PTDGGTSAEA MQVPLEEDDEMEEEEVINDENFLGKRPLDSPEVEEMP <small>SMKRPRLLSTKGD</small> SLDVVLL <small>EAR</small>
tr Q05BL5 Q05BL5_MOUSE	153	1.25e-60	PTDGGTSAEA MQVPLEEDDEMEEEEVINDENFLGKRPLDSPEVEEMP <small>SMKRPRLLSTKGD</small> SLDVVLL <small>EAR</small>
tr A2ASY0 A2ASY0_MOUSE	0	1.25e-60	MQVPLEEDDEMEEEEVINDENFLGKRPLDSPEVEEMP <small>SMKRPRLLSTKGD</small> SLDVVLL <small>EAR</small>
tr A2ASY1 A2ASY1_MOUSE	153	1.25e-60	PTDGGTSAEA MQVPLEEDDEMEEEEVINDENFLGKRPLDSPEVEEMP <small>SMKRPRLLSTKGD</small> SLDVVLL <small>EAR</small>
sp Q5HZG4 TAF3_MOUSE	153	1.25e-60	PTDGGTSAEA MQVPLEEDDEMEEEEVINDENFLGKRPLDSPEVEEMP <small>SMKRPRLLSTKGD</small> SLDVVLL <small>EAR</small>
tr G3GR60 G3GR60_CRIGR	153	2.24e-60	PTDGGTSAEA MQVPLEEDDEMEEEEVINDENFLGKRPLDSPEVEEMP <small>TMKRPRLLSTKGD</small> SLDVVLL <small>EAR</small>
tr D3ZPB7 D3ZPB7_RAT	153	2.34e-57	PTDGGTSAEA MQVPLEEDDELEEEEVINDENFLGKRPLDSPEVEEMP <small>SMKRPRLLNSKGD</small> SLDVVLL <small>EAR</small>
gi 74747393 sp Q5VWG9.1 TAF3_HUMAN	153	2.16e-53	PTDGGTSAEA MQVPLEEDDELEEEI IINDENFLGKRPLDSPEAEELP <small>AMKRPRLLSTKGD</small> TLDVVLL <small>EAR</small>
gi 82075449 sp Q5F489.1 TAF3_CHICK	153	1.62e-40	PTDGGTSAEA MQVPLEEEGDMEEDAINDENYLSKRPLESPDAEEFP <small>PMKRPKLSVSKGD</small> ALDGALEPRE
gi 82178917 sp Q5EAW9.1 TAF3_XENLA	84	4.69e-36	GGTSAEAMQV PLEEEEEEGEMEDDET <small>VNDENYLSKRPLDSPETMEMPFAKRIRLMNNKGD</small> ILDGSLEPRE



# 二级结构预测分析—garnier

```

      . 60 . 70 . 80 . 90 . 100
RYSELYGRDTPILDDVGEAFQLMGVNLHELEDYIHNIEPVTFFPHQIPSPF
helix      HHHHHHHHHHHHHHHHHHH
sheet EE      EEEE      EE      E
turns T TTTT      TT TT TT
coil      CC      C      CCCC      CC      C
      . 110 . 120 . 130 . 140 . 150
VSKNNVLQFPQPGSKDAEERKDYIPDYLPPIVSSQEEEEEEQVPTDGGTS
helix      HHHHH      HHHHHHHHH
sheet      EEE      E EEE
turns T TTT      TTT TT T      TT
coil C C CCCCC      CCCC      CCCC
      . 160 . 170 . 180 . 190 . 200
AEAMQVPLEEDDEMEEEVINDENFLGKRPLDSPEVEEMP SMKRPRLLST
helix HHHHHHHHHHHHHHHHHHHHHHHHH      HHHHHHHHHHH
sheet      EE
turns      TTTT      T
coil      CCCCC      C
      . 210 . 220 . 230 . 240 . 250
KGDSLDVWVLEAREPLSSINPQKTPPVLSVRVQDRADLAPPSPQPPMLA
helix      HHHHHHH
sheet      EEEE      EEEEEEEEE      EEEE
turns TTT      TTT TTT T      T T
coil C      CC CC      CCC CCCC
      . 260 . 270 . 280 . 290 . 300
PFAKSQLPIAKPLETKSFTPKTKTKASSPGQKTKSPKAALSPARLGSPIR
helix HHHHHH      HHHHH H H      H
sheet      E      EEEE
turns T      T TTTTT TTT T T      TTTT TT
coil CC C      C      CC C CCC      CC
      . 310 . 320 . 330 . 340 . 350
SPKTIPKEKSPGRSKSPKSPKIVAHVPQTPVRPETPNRTPSAMVVE
helix      HHHHH
sheet      E      EEEEE      E
turns TT TTTT TTTTT TT TT T      TT TT TT
coil C CC      C      C C C      CCC CC CCCC

```

```

#####
# Program: garnier
# Rundate: Thu 22 Dec 2011 00:05:53
# Commandline: garnier
# -auto
# -idc 0
# -sequence 493176
# -outfile 493230
# Report_format: tagseq
# Report_file: 493230
#####

=====
#
# Sequence: NM_027748.3_cdsid_NP_082024.2      from: 1      to: 932
# HitCount: 226
#
# DCH = 0, DCS = 0
#
# Please cite:
# Garnier, Osguthorpe and Robson (1978) J. Mol. Biol. 120:97-120
#
=====

      . 10 . 20 . 30 . 40 . 50
MCEYSRSRLLRVSAVICQALGWDSVQLSACHLLTDVLRYLQQLGRGCH
helix H H      HH HH
sheet EE      EEEEEEEEE      EE EEEE EEEEE
turns TTTT      TT TT T      TTTTT
coil C      C C      C
      . 60 . 70 . 80 . 90 . 100
RYSELYGRDTPILDDVGEAFQLMGVNLHELEDYIHNIEPVTFFPHQIPSPF
helix      HHHHHHHHHHHHHHHHHHH
sheet EE      EEEE      EE      E
turns T TTTTT      TT TT TT
coil      CC C      C      CCCC      CC      C

```



## 二级结构预测分析—garnier

```
#-----  
#  
# Residue totals: H:388   E:176   T:178   C:190  
#           percent: H: 42.4 E: 19.2 T: 19.4 C: 20.7  
#  
#-----  
  
#-----  
# Total_sequences: 1  
# Total_hitcount: 226  
#-----
```

这个蛋白的helix较多，在序列的N段相对较少，而在510~750处则非常多；超过连续5个残基都是beta sheet的区域有39~44，229~239，327~332，382~392，675~681，762~770，776~782，804~812，820~828，848~862。

# 二级结构预测分析—swiss-model

swiss-model 中的 structure assessment



BIOZENTRUM  
Universität Basel  
The Center for Molecular Life Sciences



SWISS-MODEL Workspace

Modelling

Tools

Repository

Documentation

[ myWorkspace ]

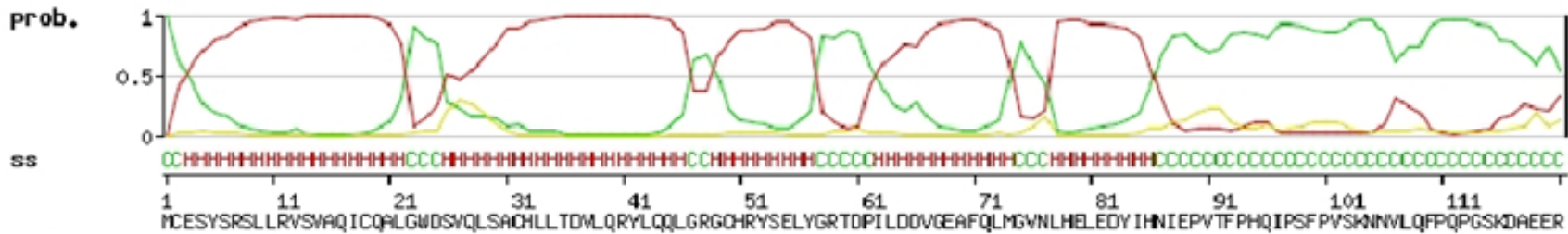
[ login ]

Workunit: P000001 TAF3

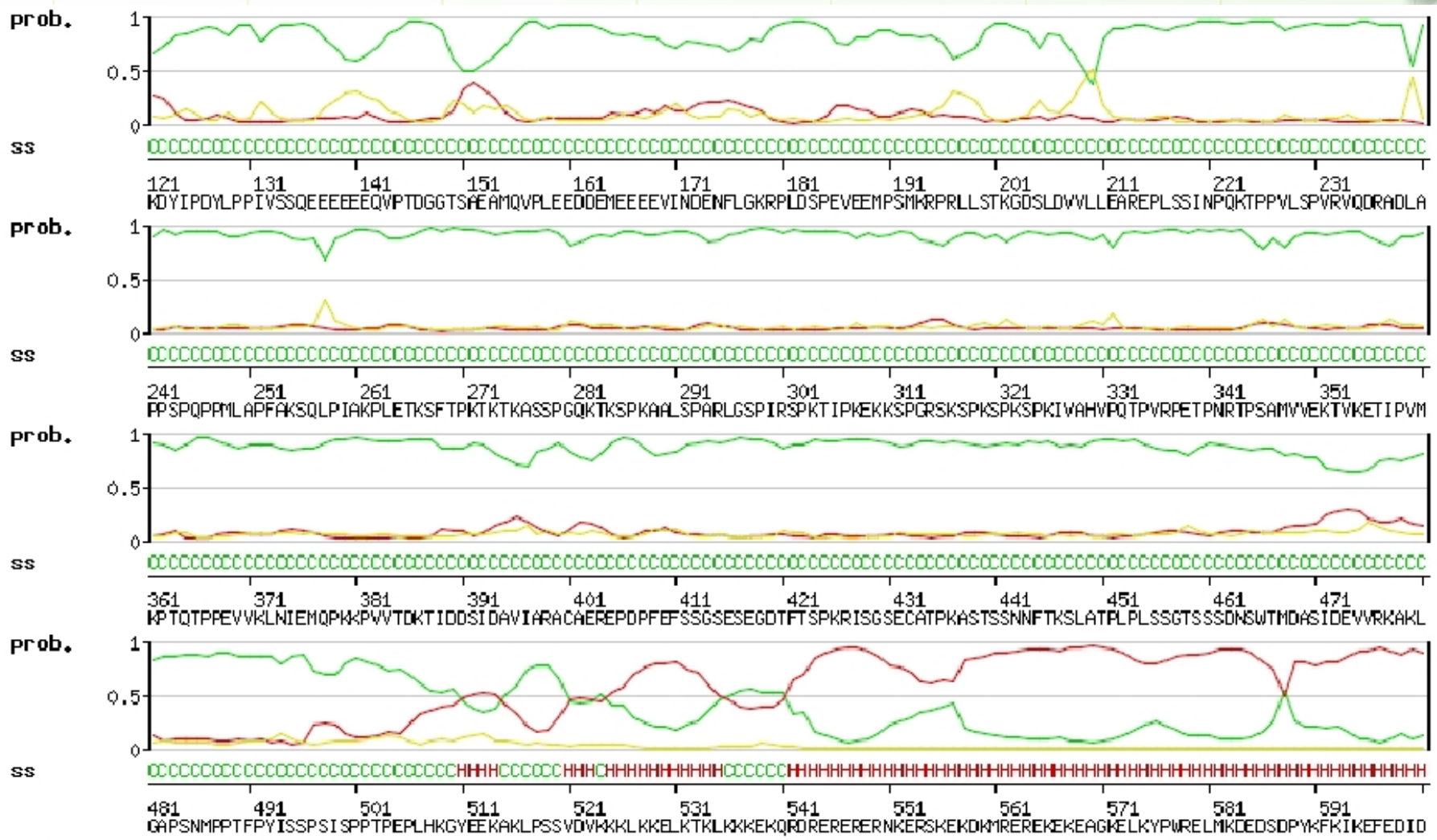
## InterproScan



- IPR001965: Zinc finger, PHD-type, Domain
- PF00628
- PS01359
- PS50016
- IPR006565: Bromodomain transcription factor, Domain
- PF07524
- IPR009072: Histone-fold, Domain
- SSF47113
- IPR011011: Zinc finger, FYVE/PHD-type, Domain
- SSF57903



# 二级结构预测分析—swiss-model





## 二级结构预测分析—swiss-model

Interpro: [+/]

Interpro Scan has finished. Here are the results:

```

IPR001965: Zinc finger, PHD-type, Domain
PF00628: 869 - 917 PHD

IPR001965: Zinc finger, PHD-type, Domain
PS01359: 870 - 914 ZF_PHD_1

IPR001965: Zinc finger, PHD-type, Domain
PS50016: 867 - 917 ZF_PHD_2

IPR006565: Bromodomain transcription factor, Domain
PF07524: 3 - 79 Bromo_IP

IPR009072: Histone-fold, Domain
SSF47113: 8 - 75Histone-fold

IPR011011: Zinc finger, FYVE/PHD-type, Domain
SSF57903: 862 - 922 FYVE_PHD_ZnF
  
```

swiss-model InterProScan结果分析:

1~90位, 是一个helix富集区,  
 3~79位可能是Bromodomain转录因子区,  
 8~75位可能是组氨酸折叠区;  
 90~500位是一段长的coli区;  
 510~730位是一段长的helix区;  
 856~861, 867~870, 908~911可能是三个beta sheet;  
 869~917或870~914或867~917可能是PHD-type的Zn指区;  
 862~922可能是FYVE/PHD-type的Zn指区。

综合分析

从这两个软件的预测结果对比来看, TAF3的510~730可能是helix; 856~861可能是beta sheet区。



# 保守位点预测分析—MEGA

获取以上的非重复蛋白序列（10条），用MEGA进行多序列比对。

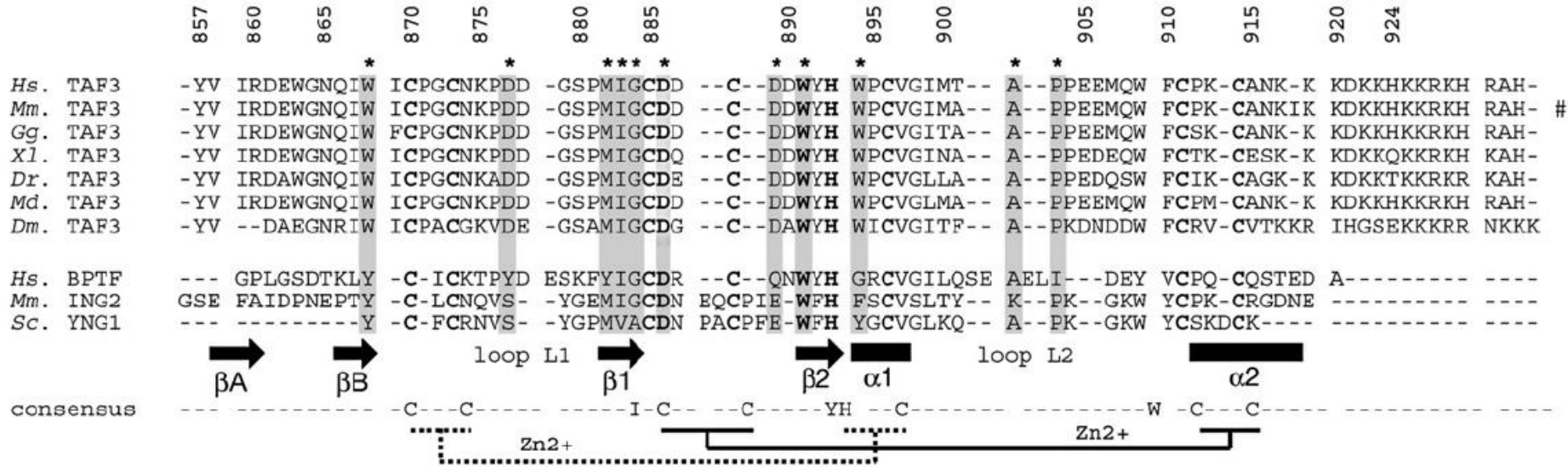
Species/Abbrv	Sequence
1. sp Q5HZG4 TAF3 MOUSE Transcription initiation factor TFIID	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
2. tr A2ASY1 A2ASY1 MOUSE TAF3 RNA polymerase II TATA box bind	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
3. tr A2ASY0 A2ASY0 MOUSE TAF3 RNA polymerase II TATA box bind	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
4. tr D3ZPB7 D3ZPB7 RAT RCG55816 OS Rattus norvegicus GN LOC10	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
5. tr G3GR60 G3GR60 CRIGR Transcription initiation factor TFIID	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
6. tr Q05BL5 Q05BL5 MOUSE Taf3 protein (Fragment) OS Mus muscu	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
7. tr Q8BKP6 Q8BKP6 MOUSE Putative uncharacterized protein (Fr	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
8. gi 74747393 sp Q5VWG9.1 TAF3 HUMAN RecName: Full Transcript	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
9. gi 82075449 sp Q5F489.1 TAF3 CHICK RecName: Full Transcript	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...
10. gi 82178917 sp Q5EAW9.1 TAF3 XENLA RecName: Full Transcrip	M...Y...LLV...VA...IC...AL...V...L...A...H...LL...V...L...Y...L...L...C...G...H...Y...S...L...Y...C...I...L...L...V...S...A...F...L...M...V...L...L...L...S...Y...I...I...I...V...F...F...

总比对长度941，位点信息如下：

	保守位点	变异位点	信息位点	单位点
位点数	537	399	162	236
百分率	57.07	42.40	17.22	25.08



# 保守位点预测分析—MEGA



Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. Cell Structure 16, 1245–1256, August 6, 2008

## 3D结构预测分析

- TAF蛋白结构分析

从PDB搜索TAF3，去掉第三个不可信结果后得到两个结果：在protein model portal中搜索Q5HZG4，得到序列Identity最高的为2K16（次高的为46%）。可见结果与PDB的搜索结果一致

ID	Name	Length		Organism
2K16	Solution structure of the free TAF3 PHD domain	75		Mouse
2K17	Solution structure of the TAF3 PHD domain in complex with a H3K4me3 peptide	H3K4me3 peptide	13	Mouse
		Transcription initiation factor TFIID subunit 3	75	

## 3D结构预测分析

- TAF蛋白结构分析

从PDB搜索TAF3，去掉第三个不可信结果后得到两个结果：在protein model portal中搜索Q5HZG4，得到序列Identity最高的为2K16（次高的为46%）。可见结果与PDB的搜索结果一致

ID	Name	Length		Organism
2K16	Solution structure of the free TAF3 PHD domain	75		Mouse
2K17	Solution structure of the TAF3 PHD domain in complex with a H3K4me3 peptide	H3K4me3 peptide	13	Mouse
		Transcription initiation factor TFIID subunit 3	75	

## 3D结构预测分析—protein model portal

PMP | Query Result:

Summary: ⓘ

Domain annotation: [ [InterPro](#) ] ↗

PHD ↗  
Bromo\_TP ↗

Colors: [Query](#) | [Sequence](#) | [Structures](#) | [Models](#) | [Domains](#) | [Selection](#)

 ⓘ

**Q5HZG4** ↗ TAF3\_MOUSE ; Transcription initiation factor TFIID subunit 3; 140 kDa TATA box-binding protein-associated factor; TBP-associated factor 3; Transcription initiation factor TFIID 140 kDa subunit; TAF(II)140; Short=TAF140; Short=TAFII-140; Short=TAFII140; Mus musculus (Mouse).

# 3D结构预测分析

PROTEIN

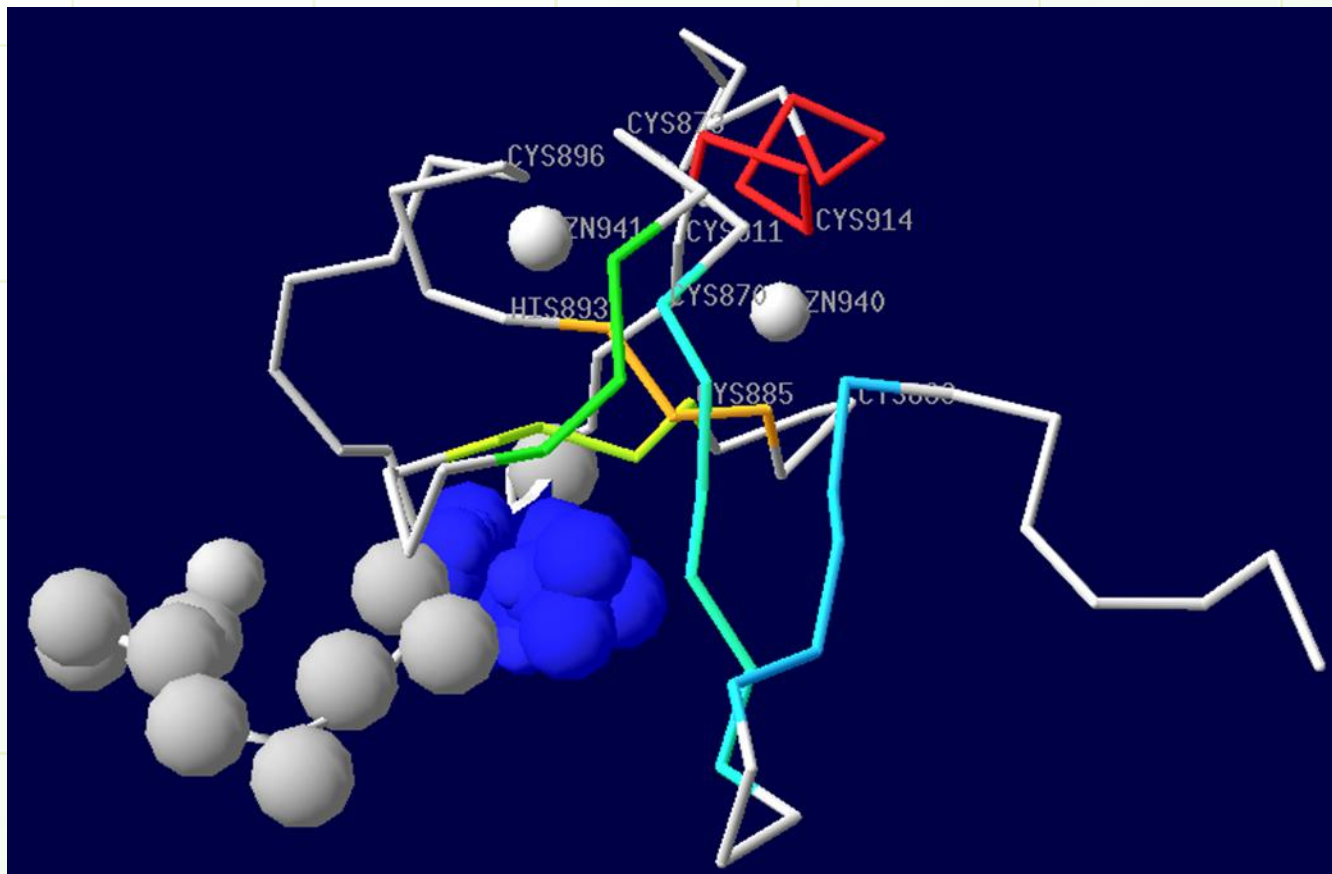
Model	Rel.	Provider	Type	Templates	%Seq id	from	to	Sel.
[Show]		MODBASE	SC	<a href="#">2k16A</a>	92%	851	924	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1e5wA</a>	46%	700	747	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">2fuiA</a>	41%	864	914	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1we9A</a>	38%	870	917	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1x4iA</a>	37%	871	931	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1w9rA</a>	32%	618	744	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1c1gA</a>	17%	516	747	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1fp0A</a>	16%	861	930	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1g8xA</a>	11%	511	747	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1deqC</a>	11%	511	751	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">2iw3A</a>	10%	723	930	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1xf1A</a>	9%	197	570	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1yv1A</a>	8%	372	746	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1cii</a>	8%	511	752	<input type="checkbox"/>
[Show]		MODBASE	SC	<a href="#">1v9dB</a>	8%	420	748	<input type="checkbox"/>

Experimental Structures:				
PDB	Title	%Seq Id	from - to	Sel.
<a href="#">2k17</a>	Solution structure of the TAF3 PHD domain in complex with a H3K4me3 peptide	100	857-924	<input type="checkbox"/>
<a href="#">2k16</a>	Solution structure of the free TAF3 PHD domain	100	857-924	<input type="checkbox"/>

## 3D结构预测分析—spdbv

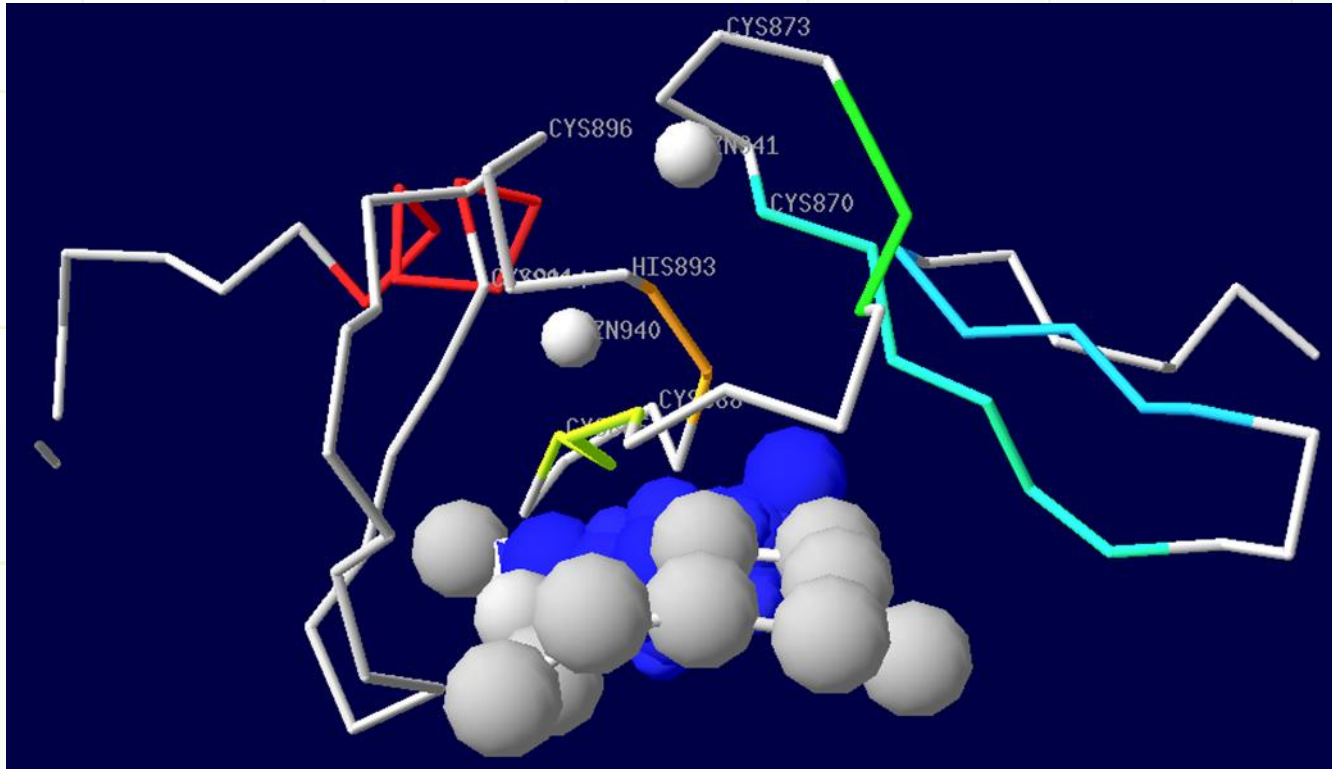
在spdb viewer 中观察2K17结构，如下图所示



中间上方的两个白色点为Zn离子，左下方的球棒结构为P链上的H3K4me3 peptide，其余的为TAFII subunit 3。Cys与His与Zn结合形成Zn指结构，继而binding在左下方的H3K4me3 peptide上。

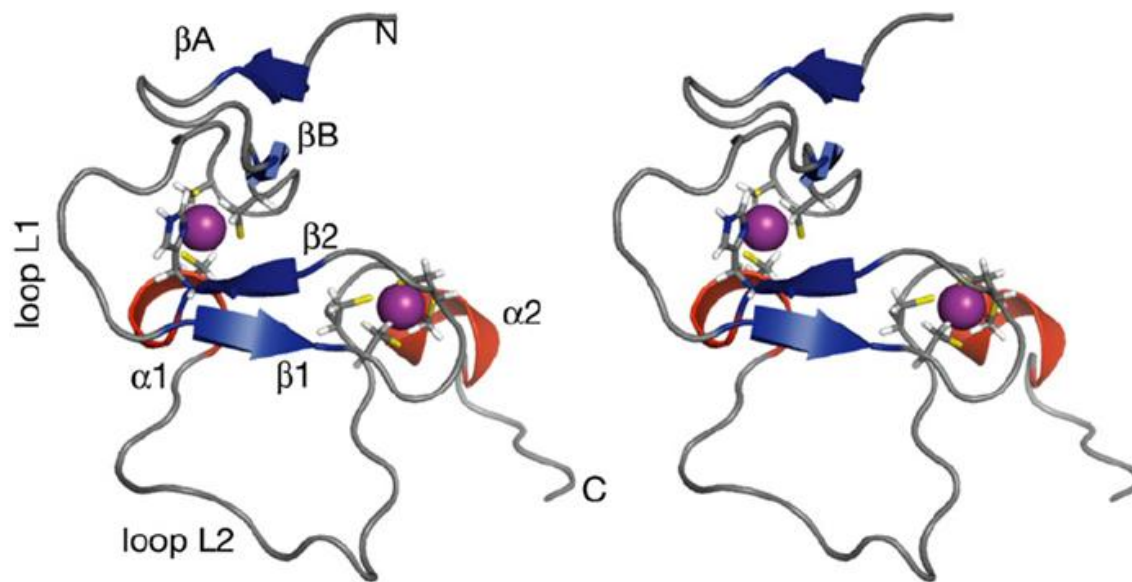


# 3D结构预测分析

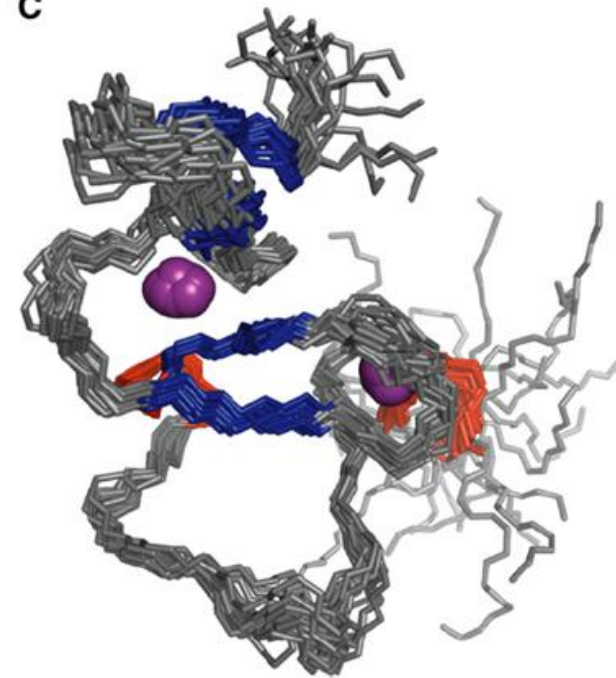


## 3D结构预测分析

B



C



Cell Structure 16, 1245–1256, August 6, 2008

## 3D结构预测分析

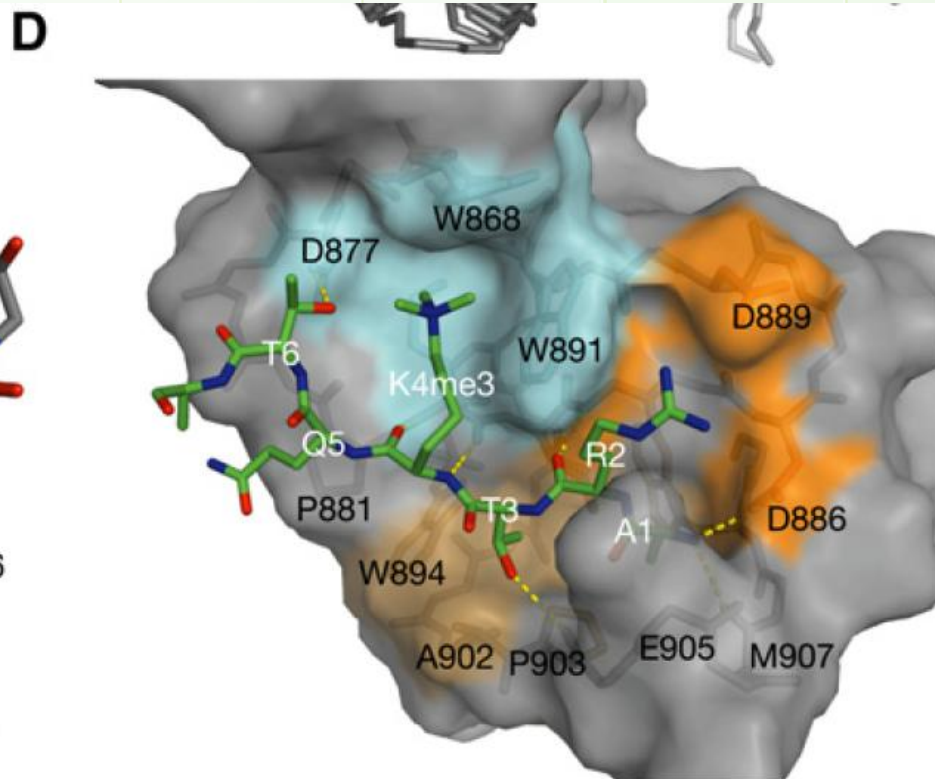
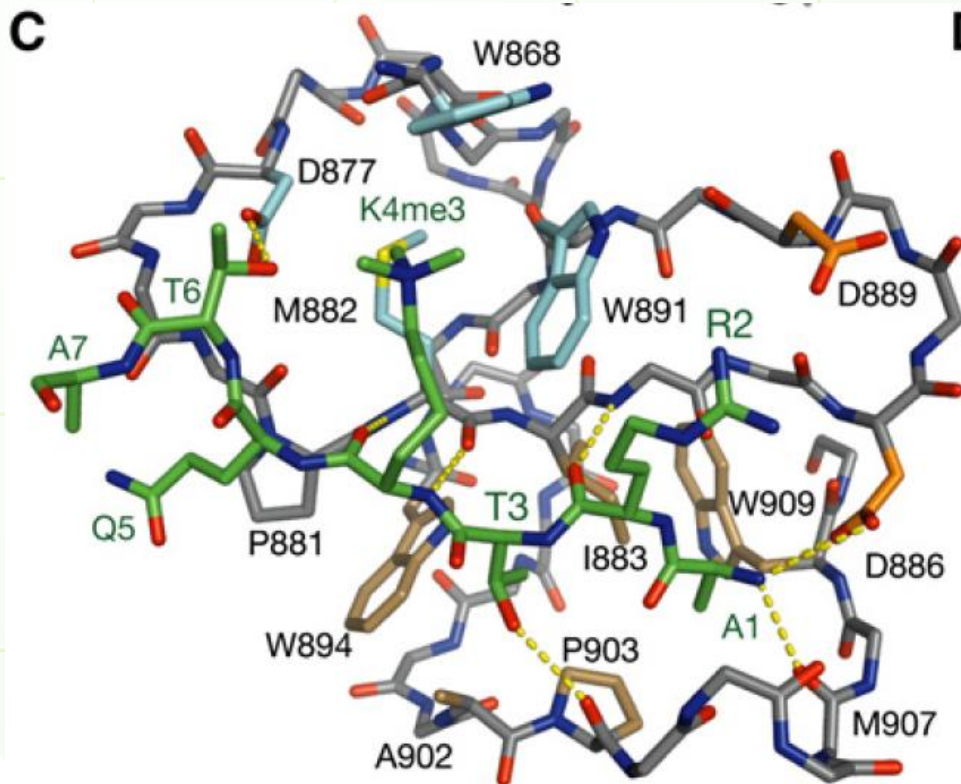
Jmol Version 12.0.41

PHD与H3K4me3的结合区域：



# 3D结构预测分析

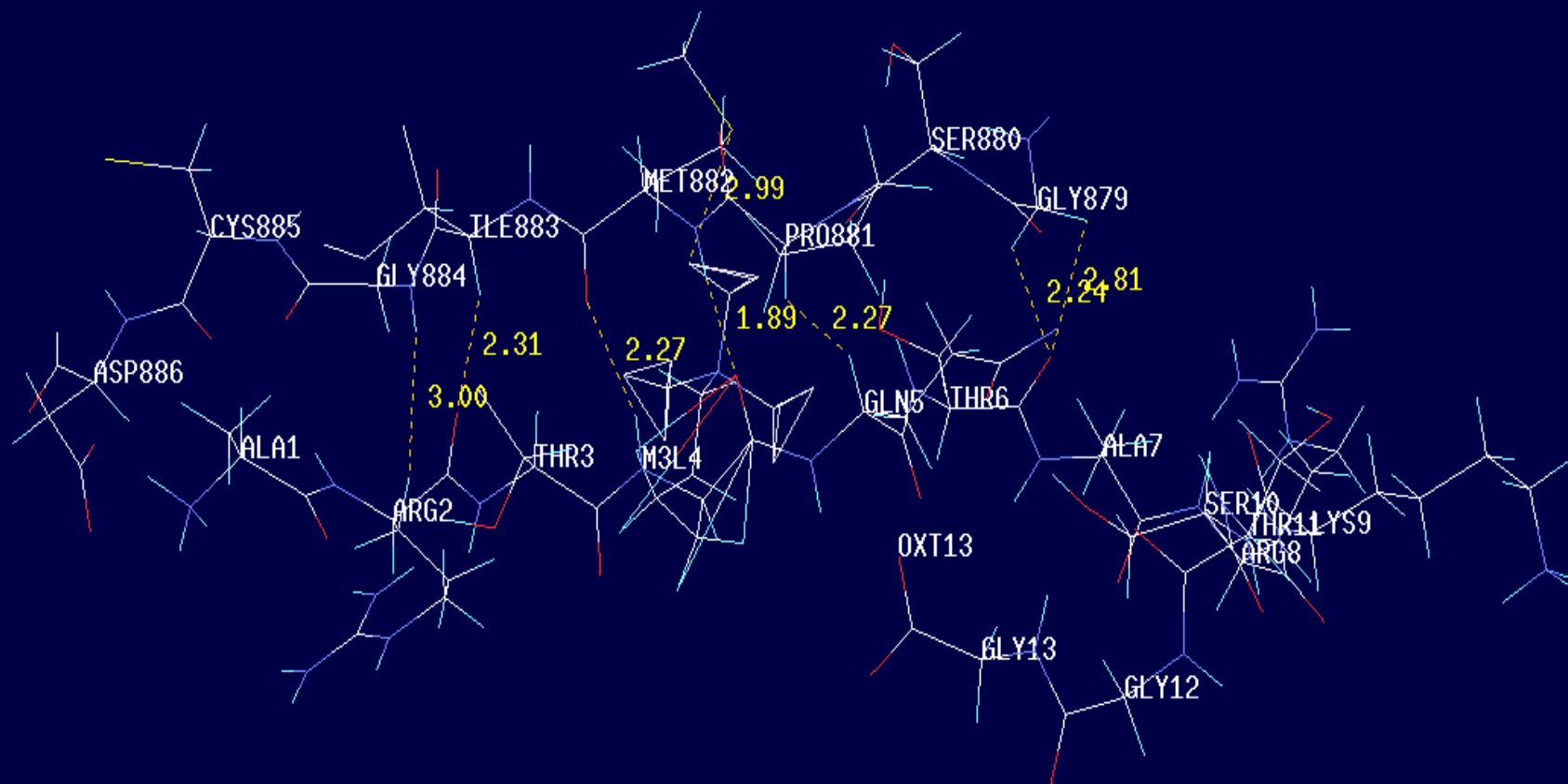
PHD与H3K4me3的结合区域:



Cell Structure 16, 1245–1256,  
August 6, 2008

## 3D结构预测分析

其部分原子距离如下图所示：



感谢各位组员的支持  
和辛苦工作!

感谢罗老师不辞辛苦  
的教诲!

L/O/G/O Thank You!

张仕坚

刘楚珺

G05

郑晓璐

彭晓红

王灿 山丹

王怡玫

G06

冯园庆