热休克蛋白70（HSP70）的生物信息学分析

组别：G13 组

组员：闵 力　　刘晓漫

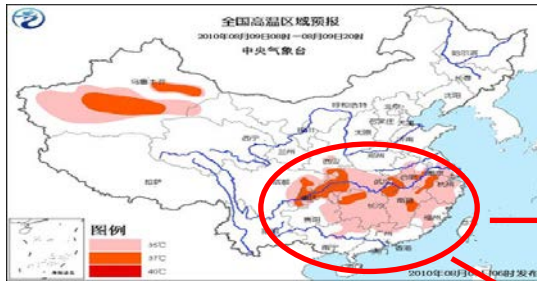　　唐 静　　饶伟丽

2014年1月25日
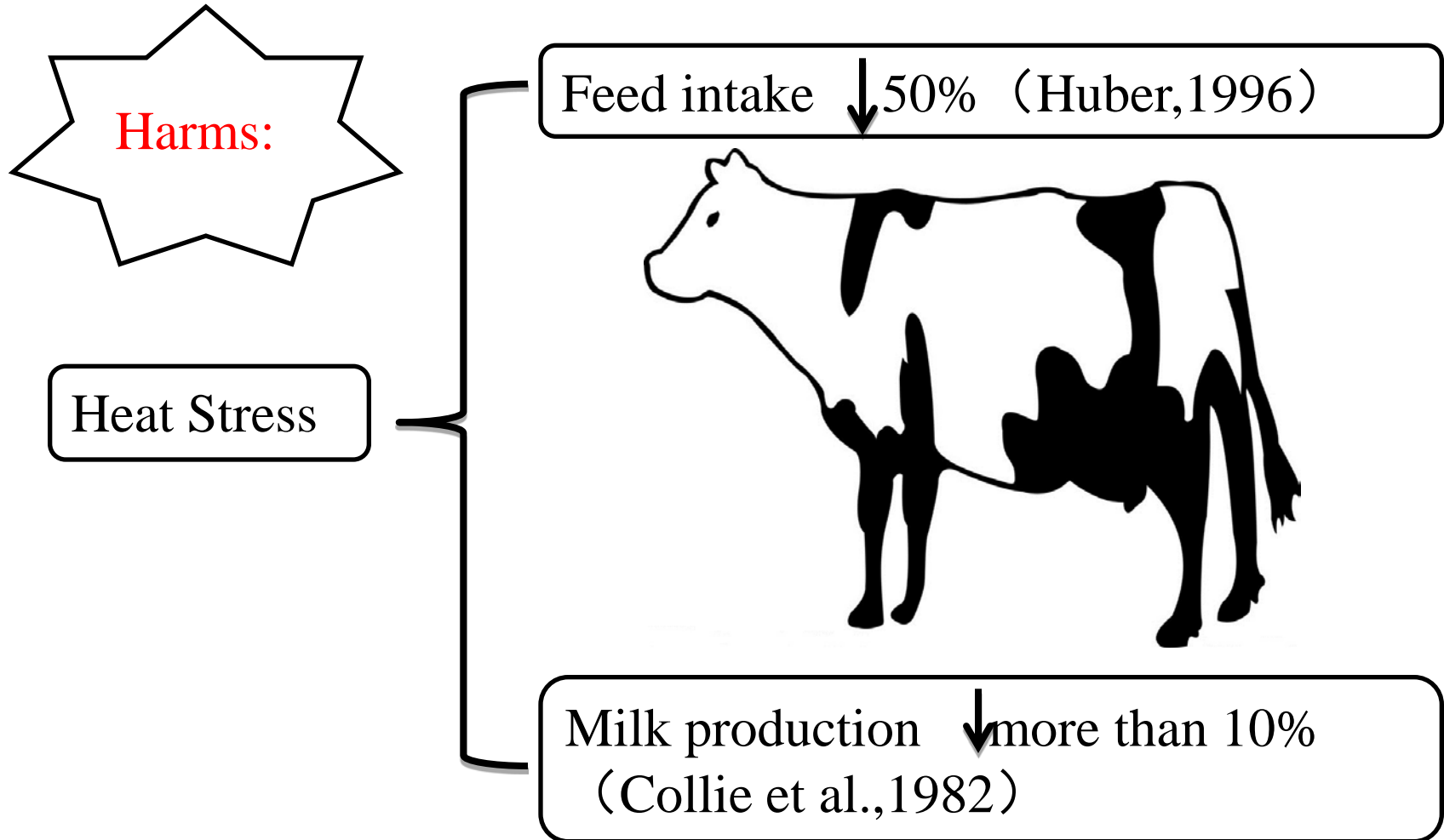
汇报内容

➢背景介绍

➢基本信息

➢蛋白序列分析

➢进化树构建

➢蛋白结构分析

➢小结及展望

Background

✓ The south of China is characterized as humid subtropical (tropical in Hainan) and is subject to extended periods of high ambient temperature and relative humidity in summer.

Background

Harms:

Heat Stress

Feed intake ↓50% （Huber,1996）

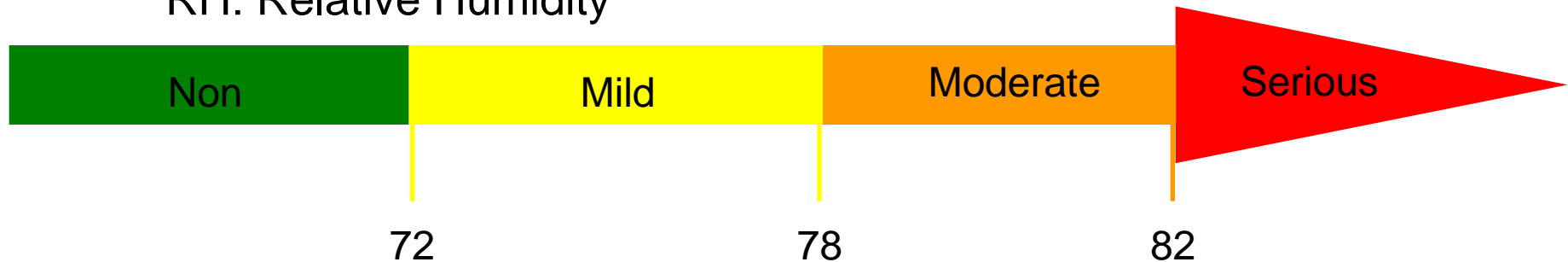Milk production ↓more than 10% （Collie et al.,1982）

Background

✓ The Temperature-Humidity Index (THI) is widely used in hot areas all over the world to assess the impact of heat stress on the livestock.

Bohmanova, et al., 2007

THI=0.81*T+（0.99*T-14.3）*RH /100+46.3

T: Temperature

RH: Relative Humidity
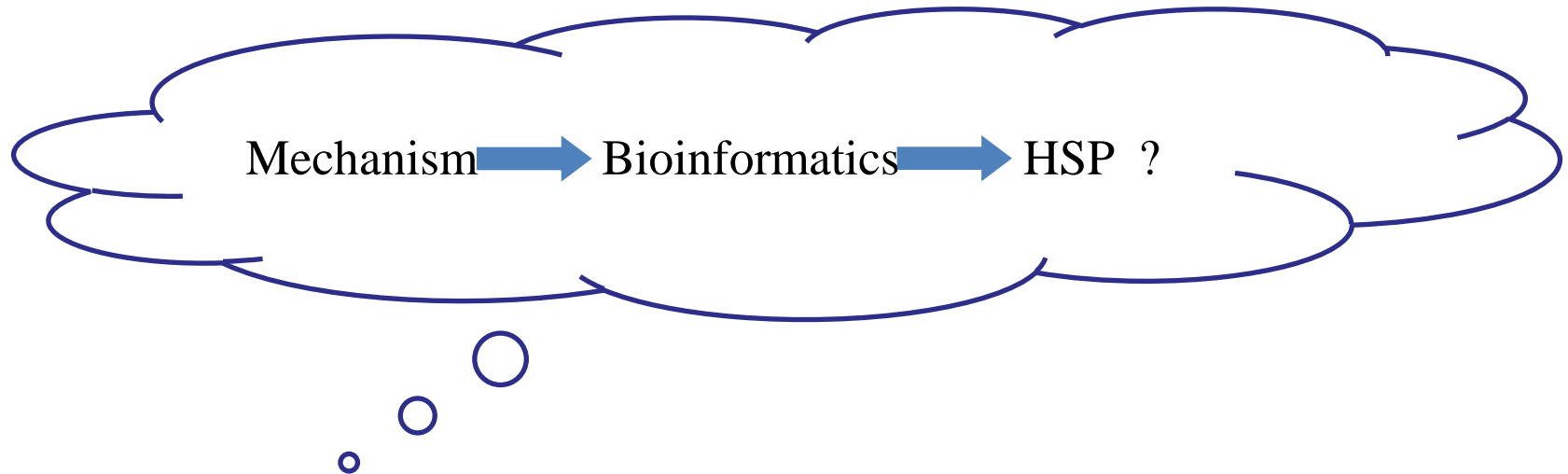
| Non | Mild | Moderate | Serious |

72　　78　　82

THI division of heat stress in dairy cow

Armstrong, 1994

Background

✓ The heat shock protein (HSP) family has long been associated with a generalized cellular stress response, particularly in terms of recognizing and chaperoning misfolded proteins.

✓ Heat Stress ➡ HSP mRNA ↑

Mechanism ➡ Bioinformatics ➡ HSP ?

生物信息学工具的初步研究和结果

◆目前，HSP70被认为是哺乳动物细胞HSPs中最重要，也是研究最为深入的HSPs家族

Uniprot数据库中对*Bovin.*HSP70.1相关信息的检索：

◆蛋白名称：Heat shock 70 kDa protein 1A

Names and origin	
Protein names	*Recommended name:* **Heat shock 70 kDa protein 1A** *Alternative name(s):* Heat shock 70 kDa protein 1 Short name=HSP70.1
Gene names	Name: **HSPA1A** Synonyms:HSP70-1
Organism	**Bos taurus (Bovine)** [Reference proteome]

◆蛋白大小：641 AA

Sequence		Length	Mass (Da)	Tools	
Q27975 [UniParc].	FASTA	641	70,259	Blast	go

Last modified December 21, 2004. Version 2.
Checksum: 6FAC1F30540E80F9

```
        10         20         30         40         50         60
MAKNMAIGID LGTTYSCVGV FQHGKVEIIA NDQGNRTTPS YVAFTDTERL IGDAAKNQVA

        70         80         90        100        110        120
LNPQNTVFDA KRLIGRKFGD PVVQSDMKHW PFRVINDGDK PKVQVSYKGE TKAFYPEEIS

       130        140        150        160        170        180
SMVLTKMKEI AEAYLGHPVT NAVITVPAYF NDSQRQATKD AGVIAGLNVL RIINEPTAAA

       190        200        210        220        230        240
IAYGLDRTGK GERNVLIFDL GGGTFDVSIL TIDDGIFEVK ATAGDTHLGG EDFDNRLVNH

       250        260        270        280        290        300
FVEEFKRKHK KDISQNKRAV RRLRTACERA KRTLSSSTQA SLEIDSLFEG IDFYTSITRA

       310        320        330        340        350        360
RFEELCSDLF RSTLEPVEKA LRDAKLDKAQ IHDLVLVGGS TRIPKVQKLL QDFFNGRDLN

       370        380        390        400        410        420
KSINPDEAVA YGAAVQAAIL MGDKSENVQD LLLLDVAPLS LGLETAGGVM TALIKRNSTI

       430        440        450        460        470        480
PTKQTQIFTT YSDNQPGVLI QVYEGERAMT RDNNLLGRFE LSGIPPAPRG VPQIEVTFDI

       490        500        510        520        530        540
DANGILNVTA TDKSTGKANK ITITNDKGRL SKEEIERMVQ EAEKYKAEDE VQRERVSAKN

       550        560        570        580        590        600
ALESYAFNMK SAVEDEGLKG KISEADKKKV LDKCQEVISW LDANTLAEKD EFEHKRKELE

       610        620        630        640
QVCNPIISRL YQGAGGPGAG GFGAQGPKGG SGSGPTIEEV D
```

牛HSP70.1蛋白序列Blast结果

物种及基因	E值	得分	Identity
牛HSP70.2	0.0	1310	639/641(99%)
牦牛HSP70.2	0.0	1304	636/641(99%)
猪HSP70.2	0.0	1301	633/641(99%)
苏门达腊猩猩HSP70.1	0.0	1297	633/641(99%)
狼HSP70.1	0.0	1296	633/641(98%)
猪HSP70.1	0.0	1249	624/641(97%)
大鼠HSP70.1	0.0	1238	619/641(97%)
小鼠HSP70.1	0.0	1228	612/641(95%)
小家鼠HSP70.2	0.0	1227	600/642(93%)
非洲绿猴HSP70.1	0.0	1196	619/641(97%)

Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
641	3249.0	639/641 (99.7%)	639/641 (99.7%)	0/641 (0.0%)

```
  1 MAKNMAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERL     50
    |||||.|||||||||||||||||||||||||||||||||||||||||||||
  1 MAKNTAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERL     50

 51 IGDAAKNQVALNPQNTVFDAKRLIGRKFGDPVVQSDMKEWPFRVINDGDK    100
    ||||||||||||||||||||||||||||||||||||||.||||||||||||
 51 IGDAAKNQVALNPQNTVFDAKRLIGRKFGDPVVQSDMKHWPFRVINDGDK    100

101 PKVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGHPVTNAVITVPAYF    150
    |||||||||||||||||||||||||||||||||||||||||||||||||||
101 PKVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGHPVTNAVITVPAYF    150

151 NDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDL    200
    |||||||||||||||||||||||||||||||||||||||||||||||||||
151 NDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDL    200

201 GGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHFVEEFKRKHK    250
    |||||||||||||||||||||||||||||||||||||||||||||||||||
201 GGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHFVEEFKRKHK    250

251 KDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRA    300
    |||||||||||||||||||||||||||||||||||||||||||||||||||
251 KDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRA    300

301 RFEELCSDLFRSTLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLL    350
    |||||||||||||||||||||||||||||||||||||||||||||||||||
301 RFEELCSDLFRSTLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLL    350
```
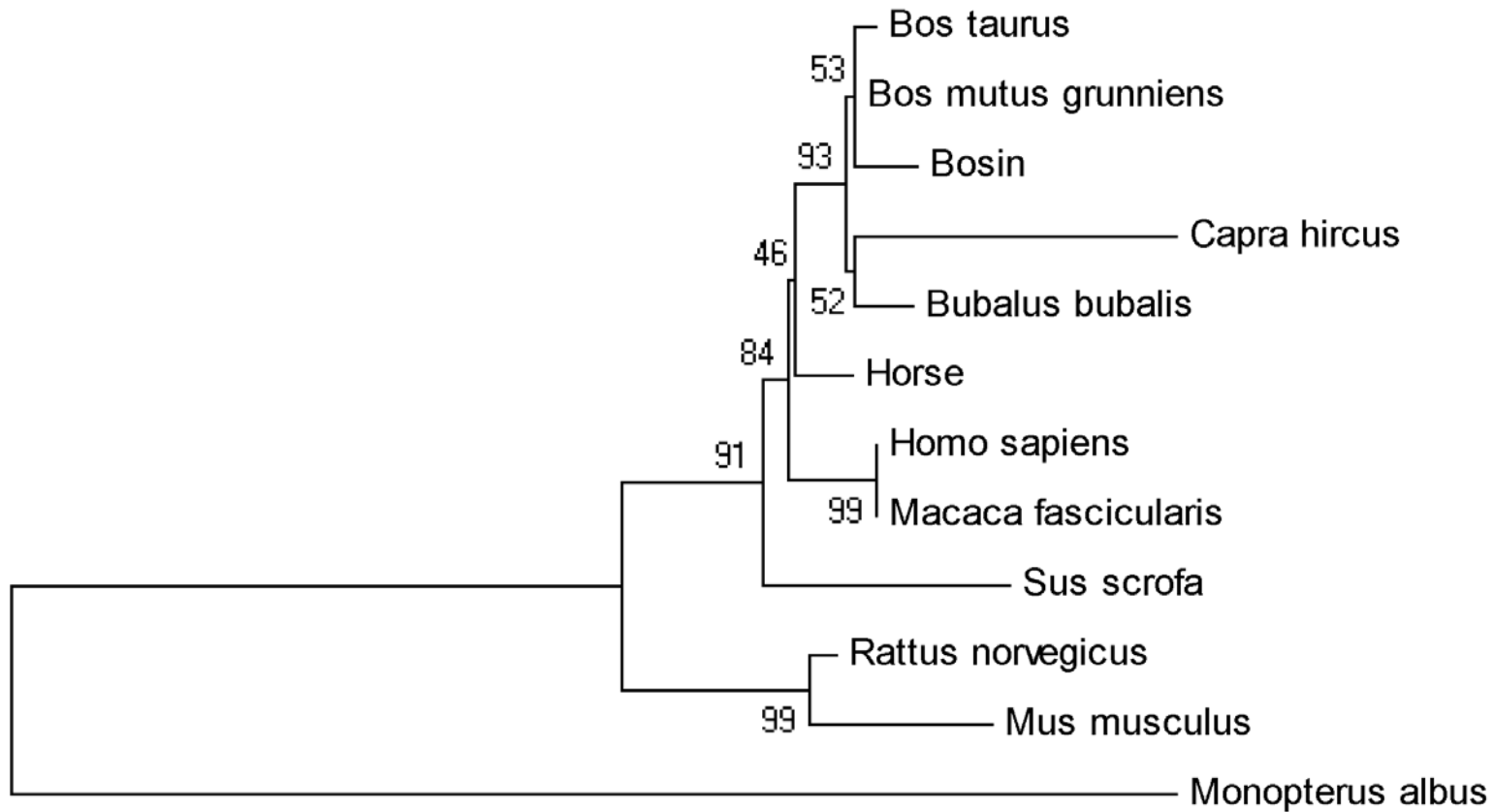
- 牛HSP70的两条蛋白序列作比对，发现第5,89位氨基酸不同，其它完全一致。

牛HSP70.1 Uniprot蛋白序列比对

物种	物种	Length	Score	Identity	Gaps
Bos grunniens	牦牛	641	3255.0	640/641 (99.8%)	0/641 (0.0%)
Bos indicus	瘤牛	641	3234.0	637/641 (99.4%)	0/641 (0.0%)
Bubalus bubalis	水牛	641	3234.0	636/641 (99.2%)	0/641 (0.0%)
Homo sapiens	人	641	3216.0	633/641 (98.8%)	0/641 (0.0%)
Equus caballus	马	642	3222.0	634/642 (98.8%)	1/642 (0.2%)
Macaca fascicularis	食蟹猴	641	3216.0	633/641 (98.8%)	0/641 (0.0%)
Sus scrofa	猪	641	3172.0	624/641 (97.3%)	0/641 (0.0%)
Capra hircus	山羊	641	3158.0	624/641 (97.3%)	0/641 (0.0%)
Rattus norvegicus	大鼠	641	3163.0	619/641 (96.6%)	0/641 (0.0%)
Mus musculus	小鼠	641	3137.0	612/641 (95.5%)	0/641 (0.0%)
Monopterus albus	黄鳝	643	2825.5	547/643 (85.1%)	5/643 (0.8%)

构建系统进化发育树

Bos taurus
53
Bos mutus grunniens
93
Bosin
Capra hircus
46
52
Bubalus bubalis
84
Horse
91
Homo sapiens
99
Macaca fascicularis
Sus scrofa
Rattus norvegicus
99
Mus musculus
Monopterus albus

H
0.002

使用ProtParam对该蛋白质进行理化性质分析：

Ala (A) 56 8.7%	Gln (Q) 26 4.1%	Leu (L) 50 7.8%	Ser (S) 34 5.3%
Arg (R) 34 5.3%	Glu (E) 48 7.5%	Lys (K) 49 7.6%	Thr (T) 41 6.4%
Asn (N) 31 4.8%	Gly (G) 54 8.4%	Met (M) 10 1.6%	Trp (W) 2 0.3%
Asp (D) 45 7.0%	His (H) 7 1.1%	Phe (F) 25 3.9%	Tyr (Y) 14 2.2%
Cys (C) 5 0.8%	Ile (I) 42 6.6%	Pro (P) 23 3.6%	Val (V) 45 7.0%

蛋白质的分子量为：70250.4D
等电点：5.55
带正点的氨基酸数目：83
带负点的氨基酸数目：93

HSP70家族蛋白分子量相近，等电点PH5.2-6.3之间，所有来源的
HSP70均富含苯丙氨酸、赖氨酸、缬氨酸、亮氨酸和脯氨酸。

使用ProtScale在线分析对该蛋白质进行亲疏水分析
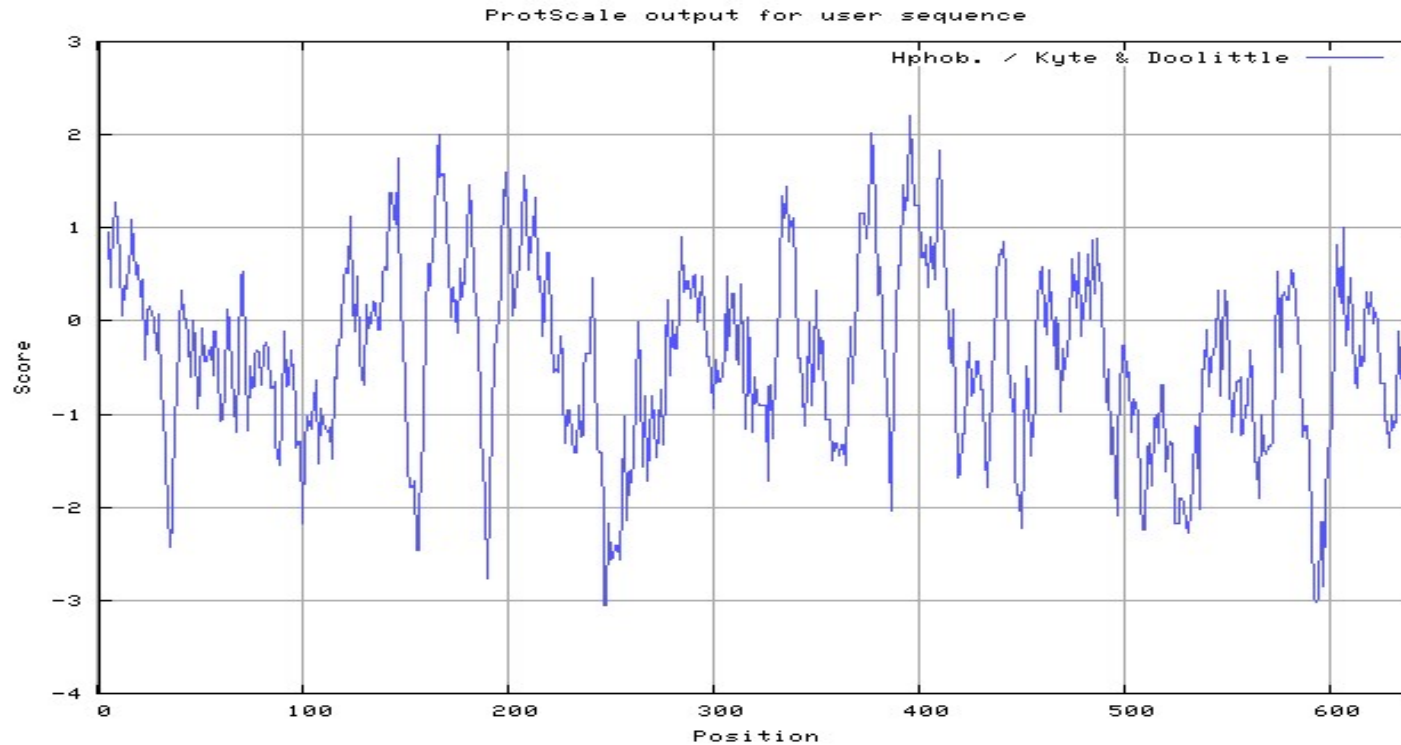
HSPA1A基因编码的蛋白质的疏水性分析结构

使用TMHMM对其进行跨膜区预测，结果如下：

WEBSEQUENCE Length: 641

WEBSEQUENCE Number of predicted TMHs: 0

WEBSEQUENCE Exp number of AAs in TMHs: 0.11338

WEBSEQUENCE Exp number, first 60 AAs: 0.03185

WEBSEQUENCE Total prob of N-in: 0.00524

WEBSEQUENCE   TMHMM2.0 outside   1   641
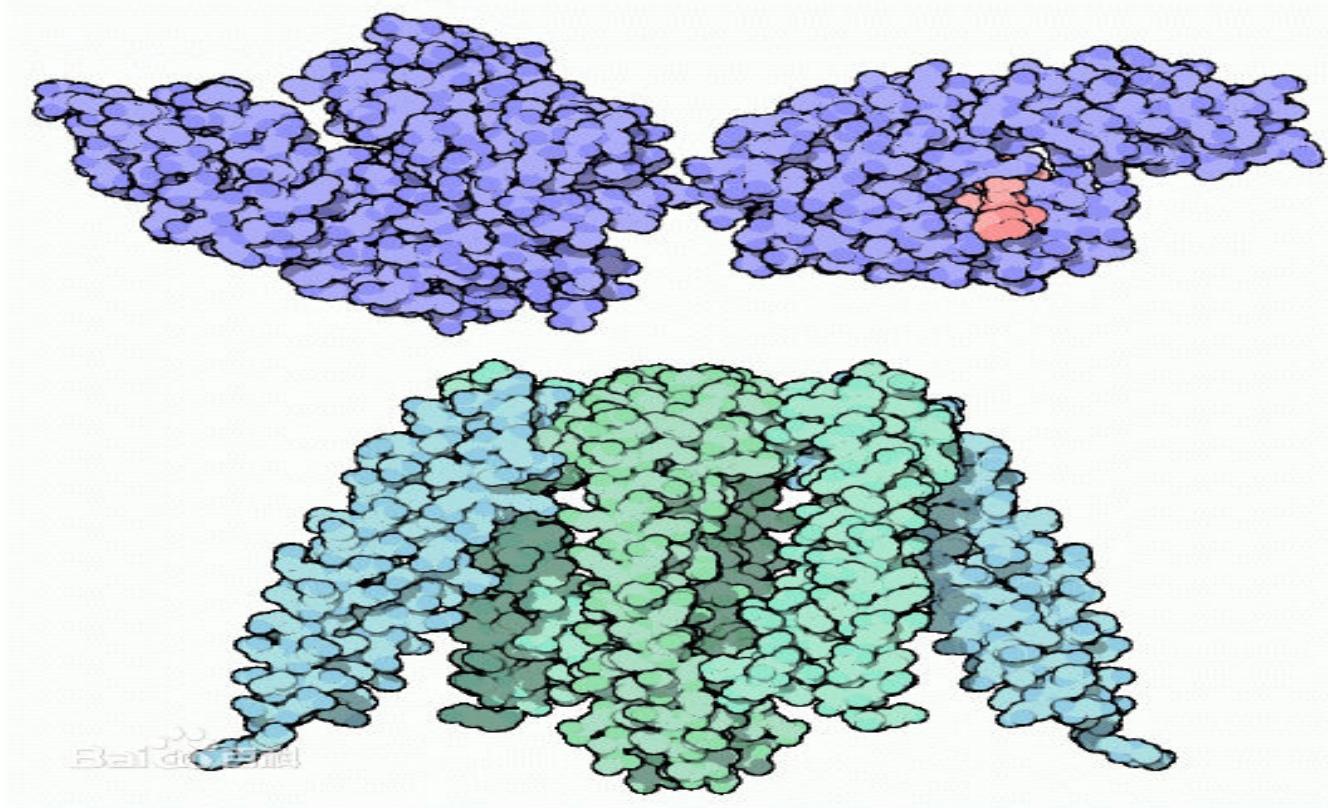
非跨膜蛋白
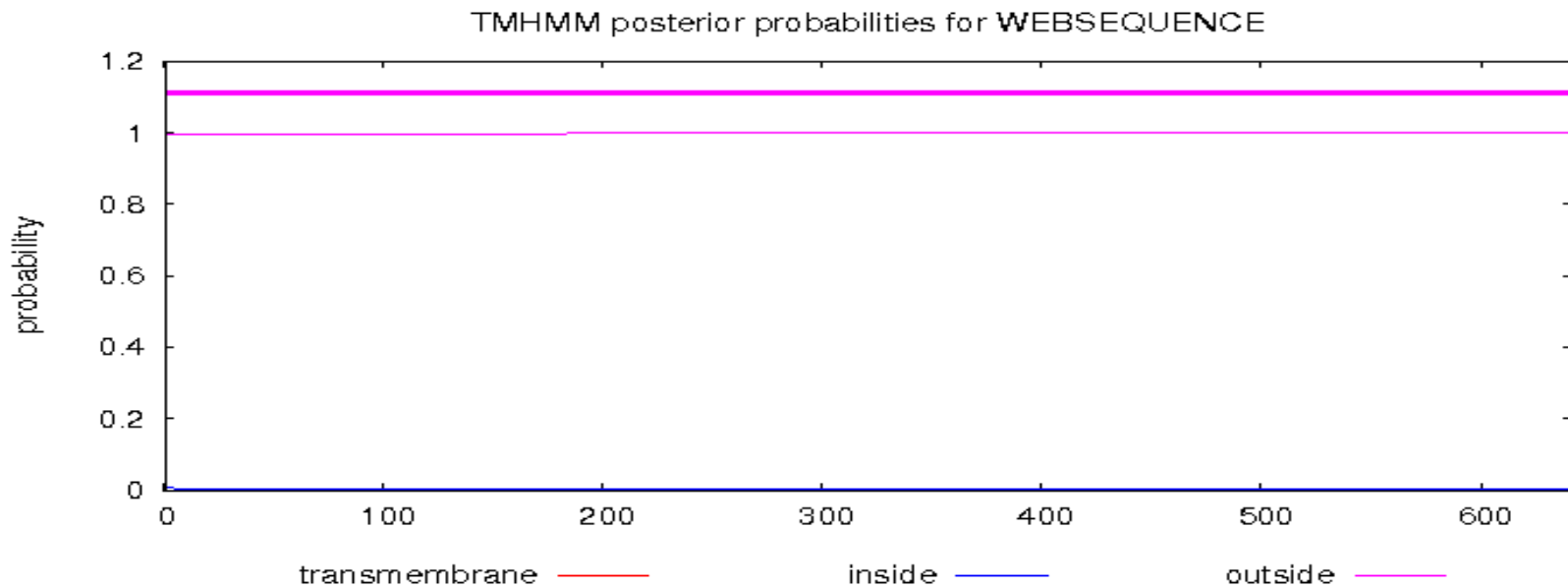
使用SignalP对该蛋白质进行信号肽预测，结果如下：

SignalP-4.1 prediction (euk networks): Sequence
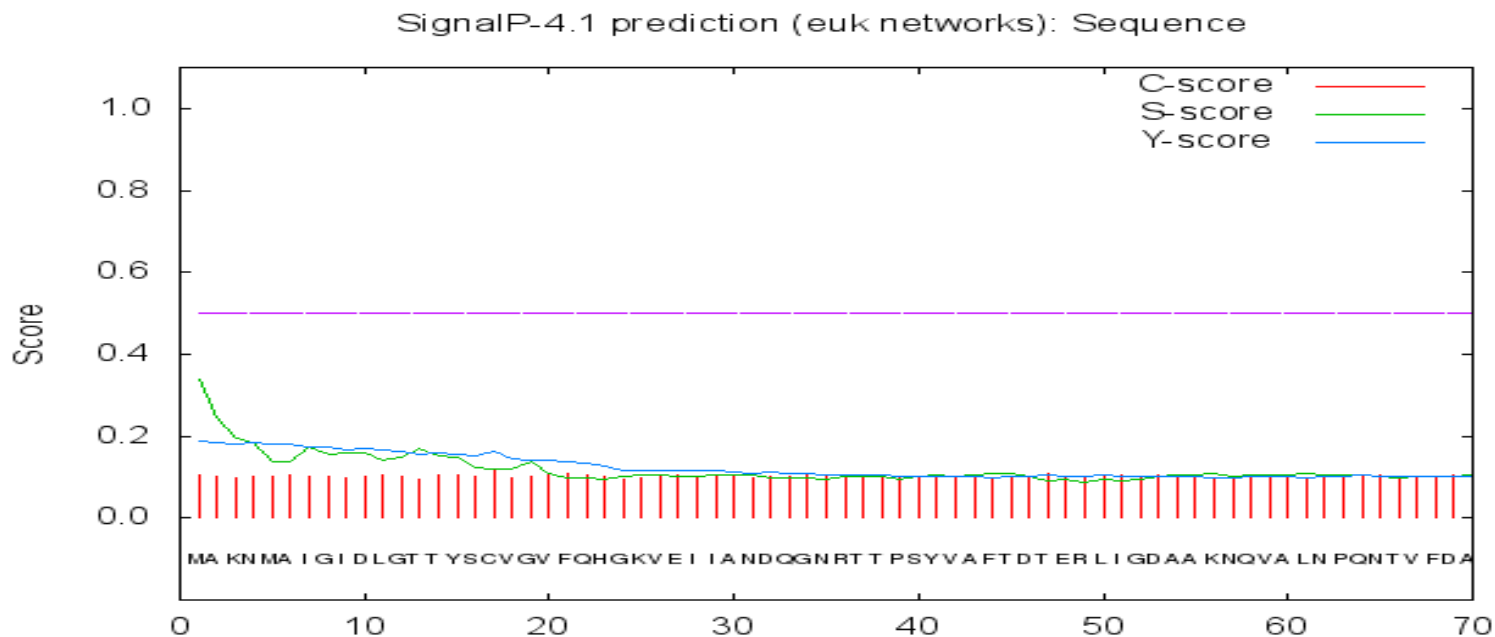
```
# Measure    Position   Value     Cutoff    signal peptide?
  max. C       17        0.119
  max. Y       11        0.168
  max. S        1        0.338
  mean S      1-10       0.188
       D       1-10      0.179    0.450     NO
Name=Sequence    SP='NO' D=0.179 D-cutoff=0.450 Networks=SignalP-noTM
```
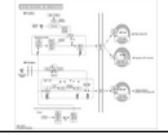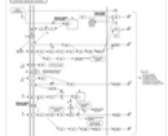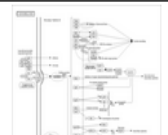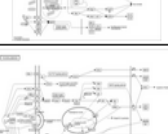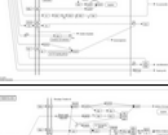
不包含信号肽

HSP70结构分析



- HSP 70由两个结构域组成：一个结构域是ATP的结合位点，并且控制ATP的结合过程
- 另一个结构域是富含碳的肽链的结合位点

KEGG相关结果

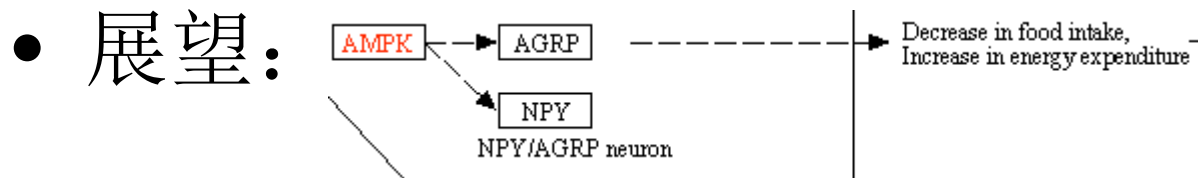| Entry	Thumbnail Image	Name	Description	Object	Legend
map04141		Protein processing in endoplasmic reticulum	The endoplasmic reticulum (ER) is a subcellular organelle where proteins are folded with the help of...	G00010 C00076 (Calcium cation) G00010 G00009 G00011 G00011 G10694 G00012 K04374 (ATF4) K04452 (DDIT3...	... WFS1 RAD23 DSK2 Png1 Ubc6/7 Ufd2 Otu1 DUB DOA1 Hsp70 (Saccharomyces cerevisiae) (Mammals) Sel1L H...
map04612		Antigen processing and presentation		K06505 (CD74) K06505 (CD74) 3108 (HLA-DMA), 3109 (HLA-DMB), 449592, 449592, 481732, 481732, 445528, ...	...MHCII MHCI MHCI MHCI TAP1/2 MHCI CANX BiP HSP90 HSP70 PA28 TNFα IFNγ Proteasome T cell receptor sign...
map04915		Estrogen signaling pathway	Estrogens are steroid hormones that regulate a plethora of physiological processes in mammals, inclu...	C00951 (Estradiol-17beta) C00951 (Estradiol-17beta) C00951 (Estradiol-17beta) C00575 (3',5'-Cyclic A...	...3K Akt eNOS pCREB DNA Shc Grb2 SOS Hsp90 FKBP52 Hsp70 PI3K-Akt signaling pathway Calcium signaling p...
map05134		Legionellosis	Legionellosis is a potentially fatal infectious disease caused by the bacterium Legionella pneumophi...	C00338 (Lipopolysaccharide) C01277 (1-Phosphatidyl-1D-myo-inositol 4-phosphate) K02406 (fliC) K12805...	...thesis Hsp genes Hbs1 Intracellular replication Hsp70 DNA Complement and coagulation cascades FlaA T...
map05145		Toxoplasmosis	Toxoplasma gondii is an obligate intracellular parasite that is prevalent worldwide. The tachyzoite ...	C06314 (Lipoxin A4) C05981 (Phosphatidylinositol-3,4,5-trisphosphate) C00076 (Calcium cation) K04729...	...PK1 PLA2 Laminin MIC3 SAG TLR11 PFTG MD2 TLR4 TgHSP70 TLR2 IFNγ IFNGR CCR5 C18 CD154 CD40 STAg DNA D...
map05152		Tuberculosis	Tuberculosis, or TB, is an infectious disease caused by Mycobacterium tuberculosis. One third of the...	C04549 (1-Phosphatidyl-1D-myo-inositol 3-phosphate) C01194 (1-Phosphatidyl-D-myo-inositol) C00338 (L...	...B β-Glucans/ mycobacterial ligands IFNγR1 HSP65 HSP70 Non-mannose-capped LAM (AraLAM) JNK p38 DNA DN...

STRING 数据库中相关结果
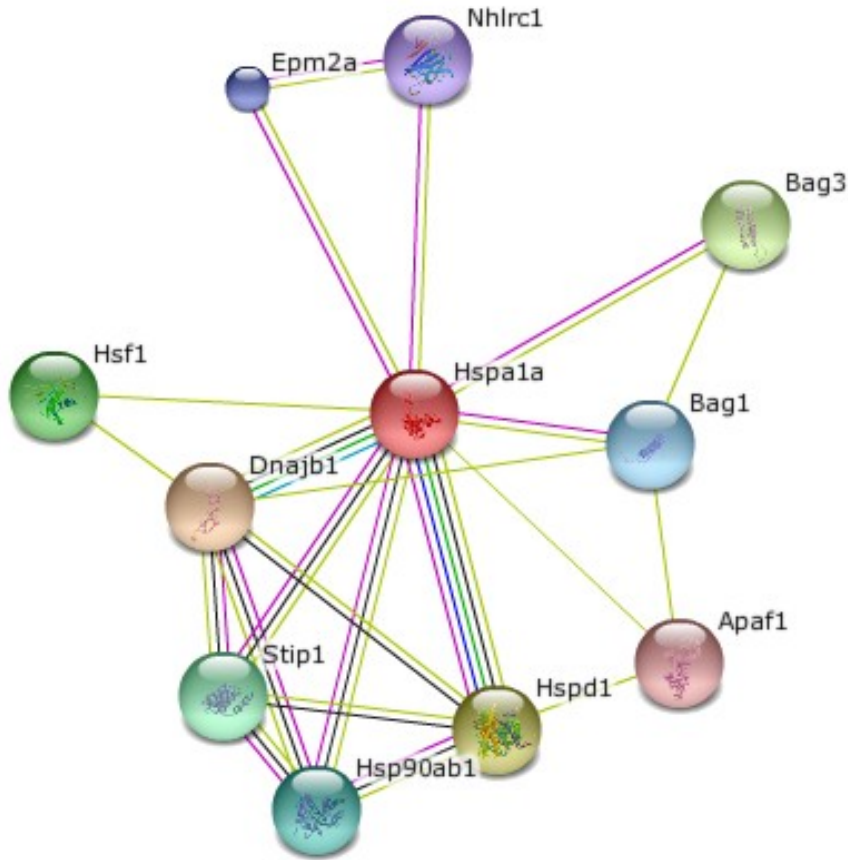
可能的相互作用蛋白：

1. Dnajb1   分值：0.993
2. Hspd1   分值：0.993
3. Bag3   分值：0.953
4. Hsf1   分值：0.950
5. Hsp90ab1 分值：0.945

小结及展望

- 现阶段：HSP的研究主要还是集中在m RNA表达水平，应激状态下的机制尚不清楚。

- 然而，通过分子信息学的初步分析，未找到热应激条件下，HSP研究的突破口。

- 展望：

致谢

- 感谢罗老师

- 感谢组员及同学们