

Bioinformatic Analysis of associated Protein HMGCR in cow

报告人：李金梅 尚利明
魏金金 张新英
第 18 组

内 容:

- 背景
- 生物信息分析
- 参考文献
- 致谢

背景

- **HMGCR**: *3-hydroxy-3-methylglutaryl-Coenzyme A reductase*;
- *This transmembrane glycoprotein is involved in the control of cholesterol biosynthesis. It is the rate-limiting enzyme of sterol biosynthesis.*

胆固醇的生物合成过程

HMGCR

乙酰CoA → 甲羟戊酸 → 二甲烯丙
基焦磷酸 → 鲨烯 → 胆固醇

人体血液中过高的胆固醇会导致与心血管相关的一些疾病。美国LRCP (lipid research clinic program) 研究结果表明, 通过调节膳食可改变人体内胆固醇的水平, 血浆胆固醇含量每降低1%, 冠心病的发病率可减少2%。

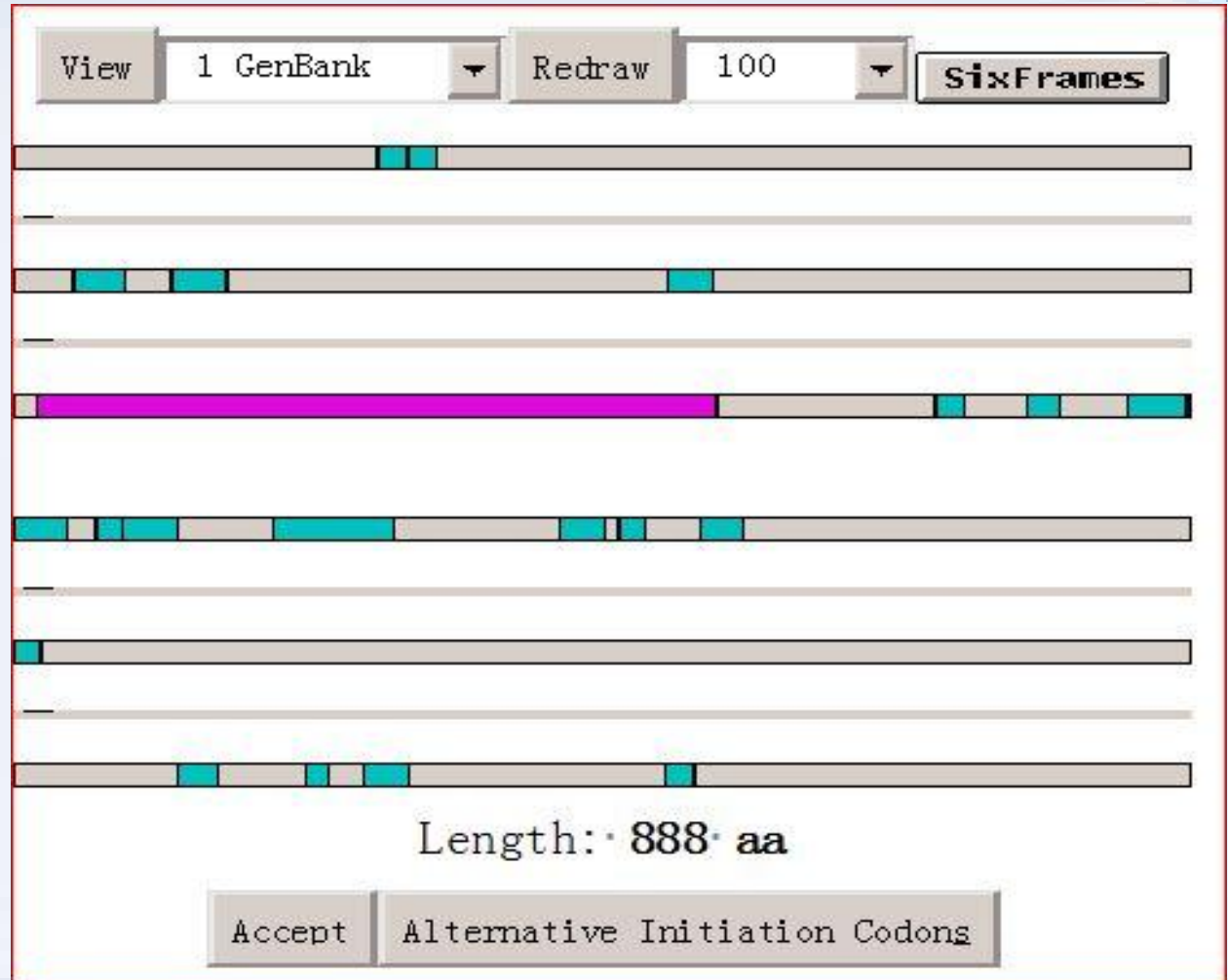
因此, 降低畜禽产品中的胆固醇含量是营养学家亟待解决的问题。畜禽HMGCR 及其调控机制的研究已成为畜禽胆固醇代谢调控领域的热点。

本研究利用生物信息学的方法，对牛（*bos taurus*）HMGCR 的基因及相应氨基酸序列的理化性质、结构特征、生化功能及系统演化关系等进行预测和分析，以期为深入开展畜禽HMGCR的酶学特性、胆固醇生物合成及调控的分子机制研究提供理论依据。

生物信息学分析

1. 用ORF

Finder在线软件查找和翻译牛HMGCR的mRNA序列的开放阅读框，预测出**6种**情况，其中第三种（图中紫色）为正确。



由Genbank中检索出牛的HMGCR mRNA全长为4618bp。其中包括5' - 和3' -UTR及1个开放阅读框，起始密码子为ATG，终止密码子为TGA，编码区为96—2762。3' -UTR中还包含一个polyA尾巴。

2. 应用**ProtParam**软件对牛的**HMGCR**进行氨基酸序列组成、蛋白质相对分子质量、疏水性、等电点等理化特性分析。

<http://web.expasy.org/cgi-bin/protparam/protparam>

分析结果

Number of amino acids: 888

Molecular weight: 97679.3

Theoretical pI: 6.27

Instability index: 51.30（大于40的为不稳定）。

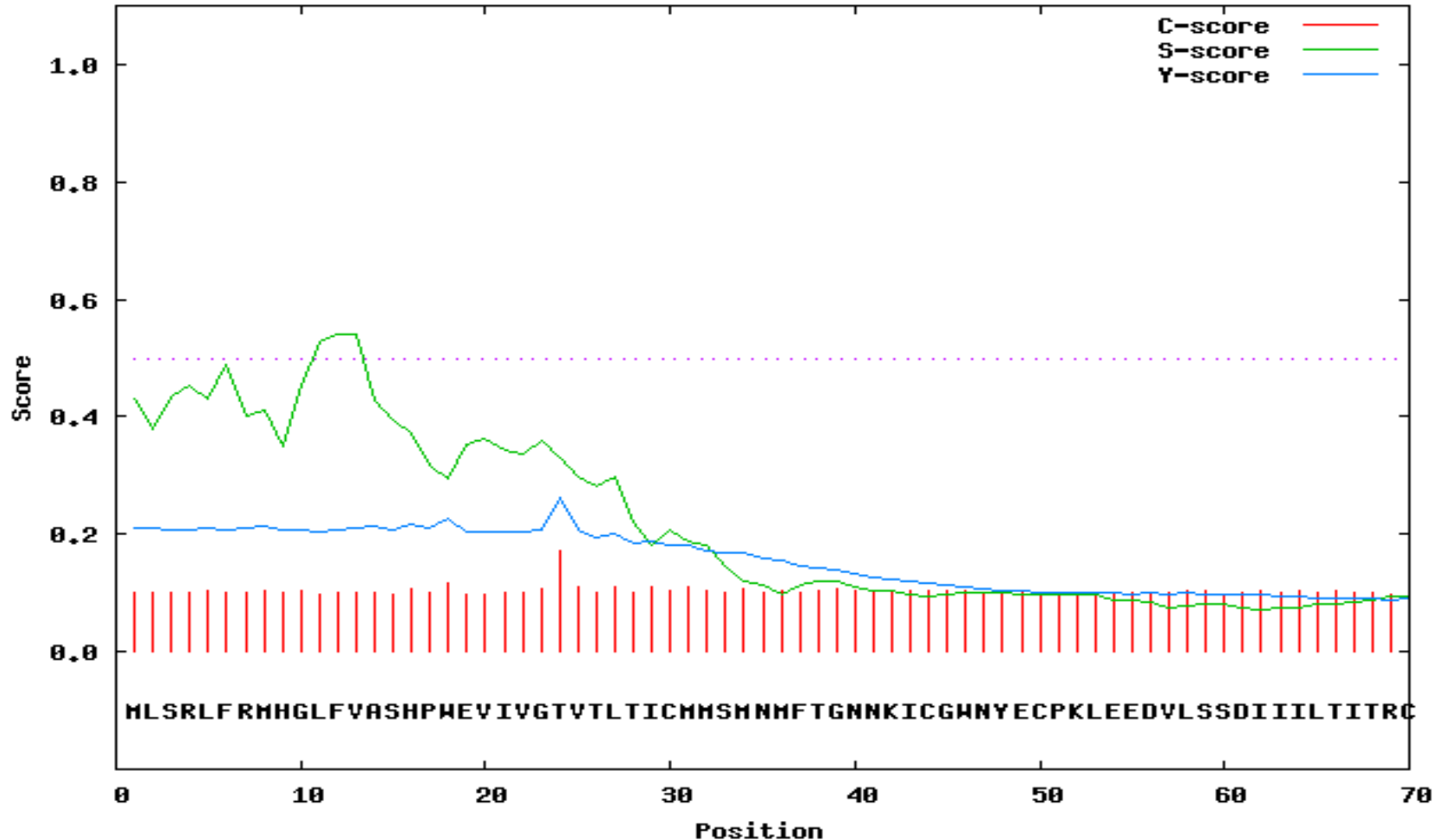
表 1 不同动物 *HMGR* 基因的核酸序列及对应氨基酸序列的组成成分及理化性质分析

指标	猪	牛	鸡
基因全长/bp	2 864	4 618	3 387
开放阅读框/bp	2 658	2 667	2 628
起始位点及密码子/bp	104,ATG	96,ATG	10,ATG
终止位点及密码子/bp	2 761,TGA	2 762,TGA	2 637,TGA
5'/3'非翻译区长度	103/103	95/1856	9/650
推导氨基酸残基数	885	888	875
分子式	C4295H6906N11600I277S60	C4309H6942N11740I281S62	C4219H6807N11330I263S57
分子量 /ku	97 150.6	97 679.3	95 439.7
理论等电点	6.14	6.27	6.22
含量最丰富的氨基酸	Leu (10.2%)、Ala (7.9%)、Ser (7.9%)、Val(7.7%)、Glu(7.0%)	Leu (10.6%)、Ala (7.9%)、Ser (7.7%)、Val(7.7%)、Glu(7.2%)	Leu (10.4%)、Ala (8.7%)、Ser (7.7%)、Val(7.7%)、Glu(7.7%)
负电/正电氨基酸比例	90/83	91/85	87/82
亲水性平均系数	0.095	0.070	0.102
蛋白质不稳定性指数	50.7,属于不稳定蛋白	51.3,属于不稳定蛋白	53.79,属于不稳定蛋白

3. 使用SignalP 3.0 在线软件进行酶蛋白信号肽的分析

<http://www.cbs.dtu.dk/services/SignalP/>

SignalP-4.0 prediction (euk networks): Sequence



Measure Position Value Cutoff signal
peptide?

max. C 24 0.173

max. Y 24 0.264

max. S 13 0.542

Name=Sequence SP='NO' D=0.322 D-
cutoff=0.500 Networks=SignalP-TM

从图中可以看出，该蛋白没有信号肽。

4.跨膜结构域预测

在线软件完成<http://www.cbs.dtu.dk/services/TMHMM/>

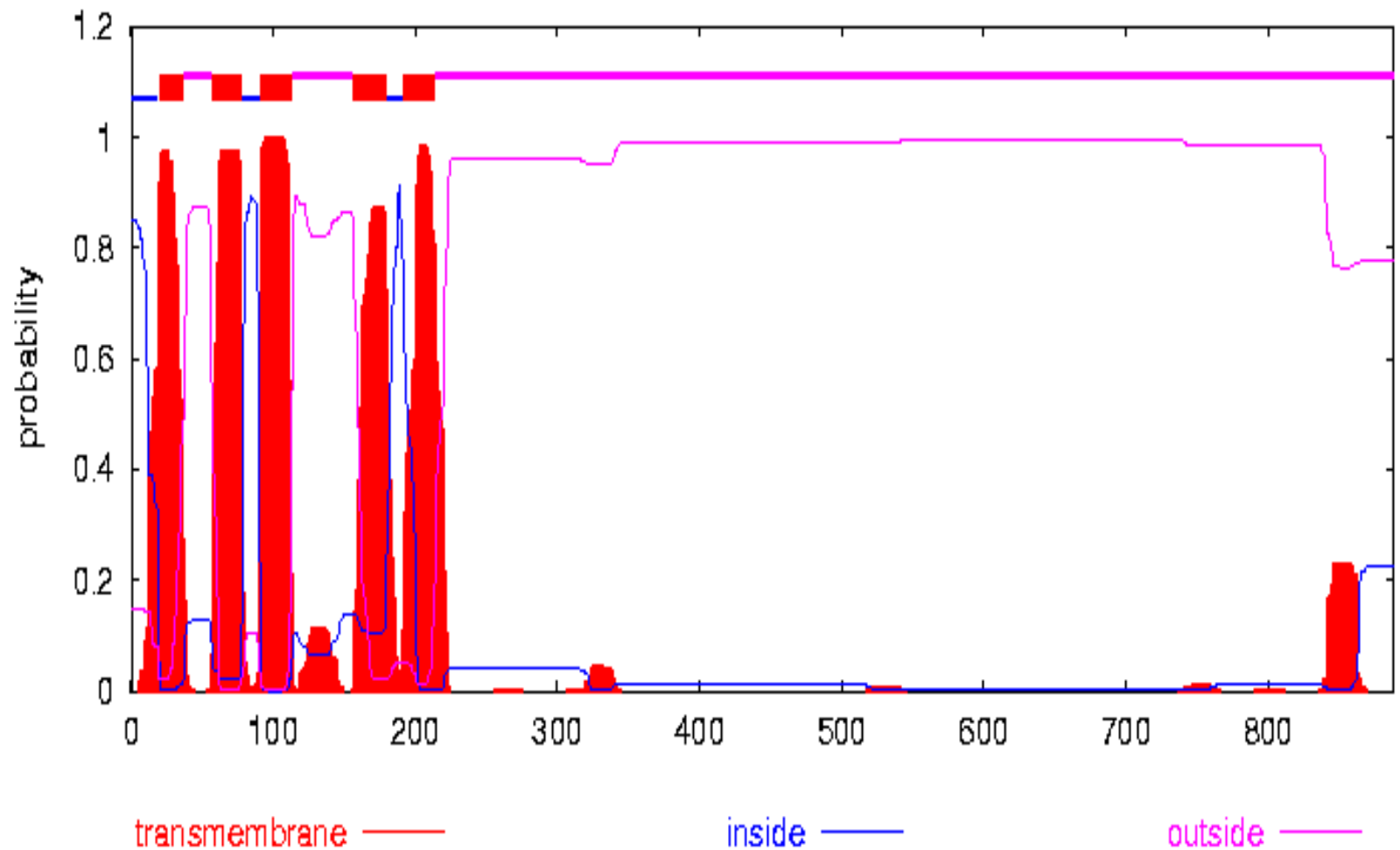
TMHMM result:

- # Sequence Length: 888**
- # Sequence Number of predicted TMHs: 5**
- # Sequence Exp number of AAs in TMHs: 113.52092**
- # Sequence Exp number, first 60 AAs: 22.72743**
- # Sequence Total prob of N-in: 0.85176**
- # Sequence POSSIBLE N-term signal sequence**

Sequence POSSIBLE N-term signal sequence

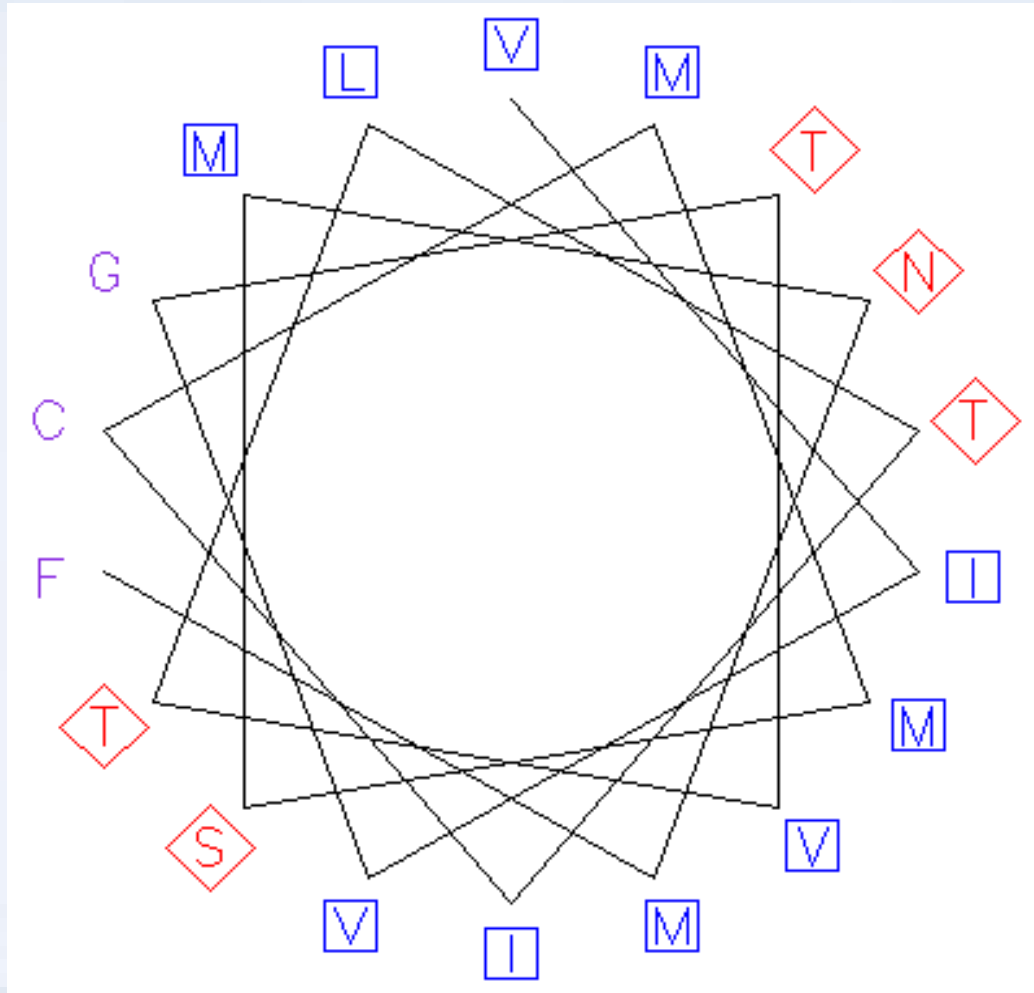
Sequence	TMHMM2.0	inside	·····	1····	19
Sequence	TMHMM2.0	TMhelix	····	20····	37
Sequence	TMHMM2.0	outside	····	38····	56
Sequence	TMHMM2.0	TMhelix	····	57····	78
Sequence	TMHMM2.0	inside	····	79····	90
Sequence	TMHMM2.0	TMhelix	····	91···	113
Sequence	TMHMM2.0	outside	···	114···	156
Sequence	TMHMM2.0	TMhelix	···	157···	179
Sequence	TMHMM2.0	inside	···	180···	191
Sequence	TMHMM2.0	TMhelix	···	192···	214
Sequence	TMHMM2.0	outside	···	215···	888

TMHMM posterior probabilities for Sequence



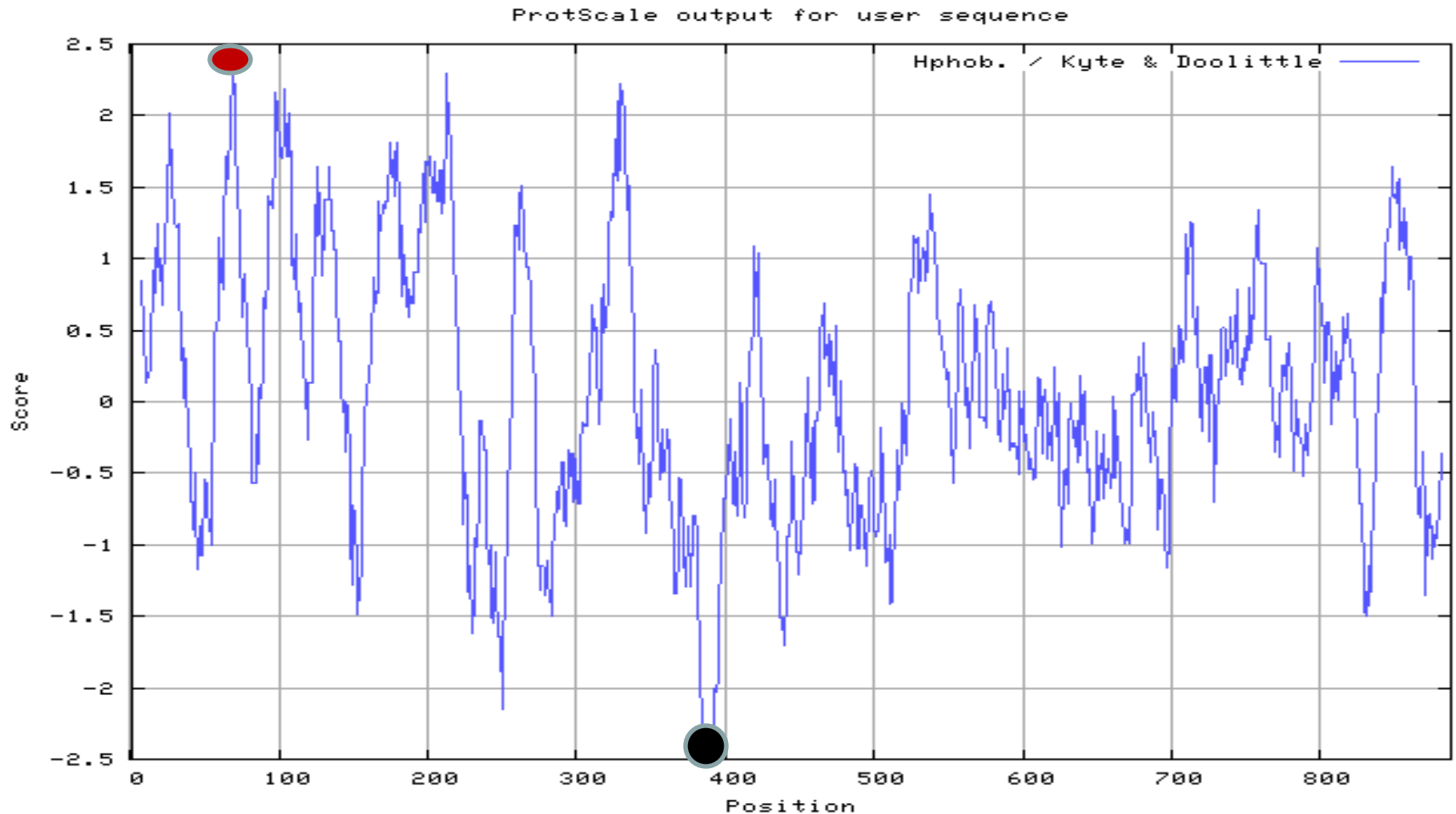
用**TMHMM 2.0 Server** 对牛**HMGCR** 氨基酸序列的跨膜结构域进行预测，结果表明牛**HMGCR**整条肽链横跨膜内外，存在**5**个跨膜结构域，分别位于**V20- F37**、**V57- F78**、**I91- F113**、**I157- I179**、**I192- A214**。

用peelwheel对第一个跨膜区绘制螺旋轮



5. 利用ProtScale分析其疏水性

<http://web.expasy.org/protscale/>



亲水性 MIN: -2.413, MAX: 2.413

6. 使用TargetP和PSORT II server工具预测HMGCR的亚细胞定位

✓ TargetP预测结果

```
### targetp v1.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions included.  
Using NON-PLANT networks.  
  
Name ..... Len ..... mTP ..... SP ..... other ..... Loc ..... RC ..... TPlen  
-----  
-  
Sequence ..... 888 ..... 0.288 ..... 0.656 ..... 0.076 ..... S ..... 4 ..... 23  
-----  
cutoff ..... 0.000 ..... 0.000 ..... 0.000
```

✓ PSORT II server工具预测亚细胞定位

65.2% : Plasma membrane 质膜

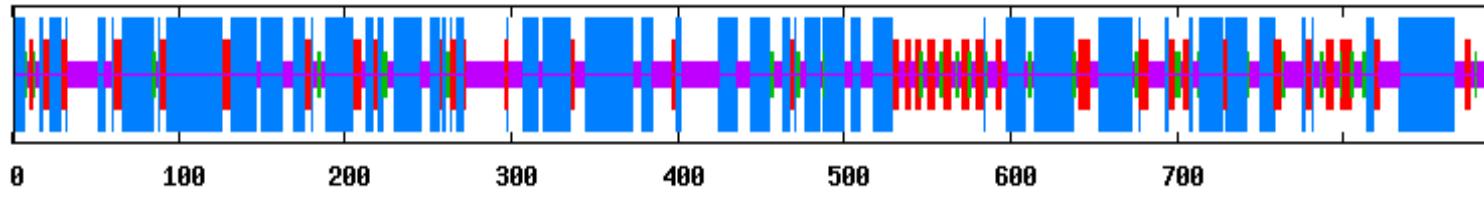
30.4%: endoplasmic reticulum 内
质网

4.3% : mitochondrial 线粒体

由结果可知，**HMGCR**定位于膜上，不存在信号肽，故推测其属于非分泌性蛋白。

7.蛋白的功能位点分析

通过**Predict Protein**对**HMGCR**氨基酸序列中可能存在的功能位点进行预测，
发现：



SOPMA:

Alpha helix (Hh) : 447 is 50.34%

Extended strand (Ee) : 127 is 14.30%

Beta turn (Tt) : 49 is 5.52%

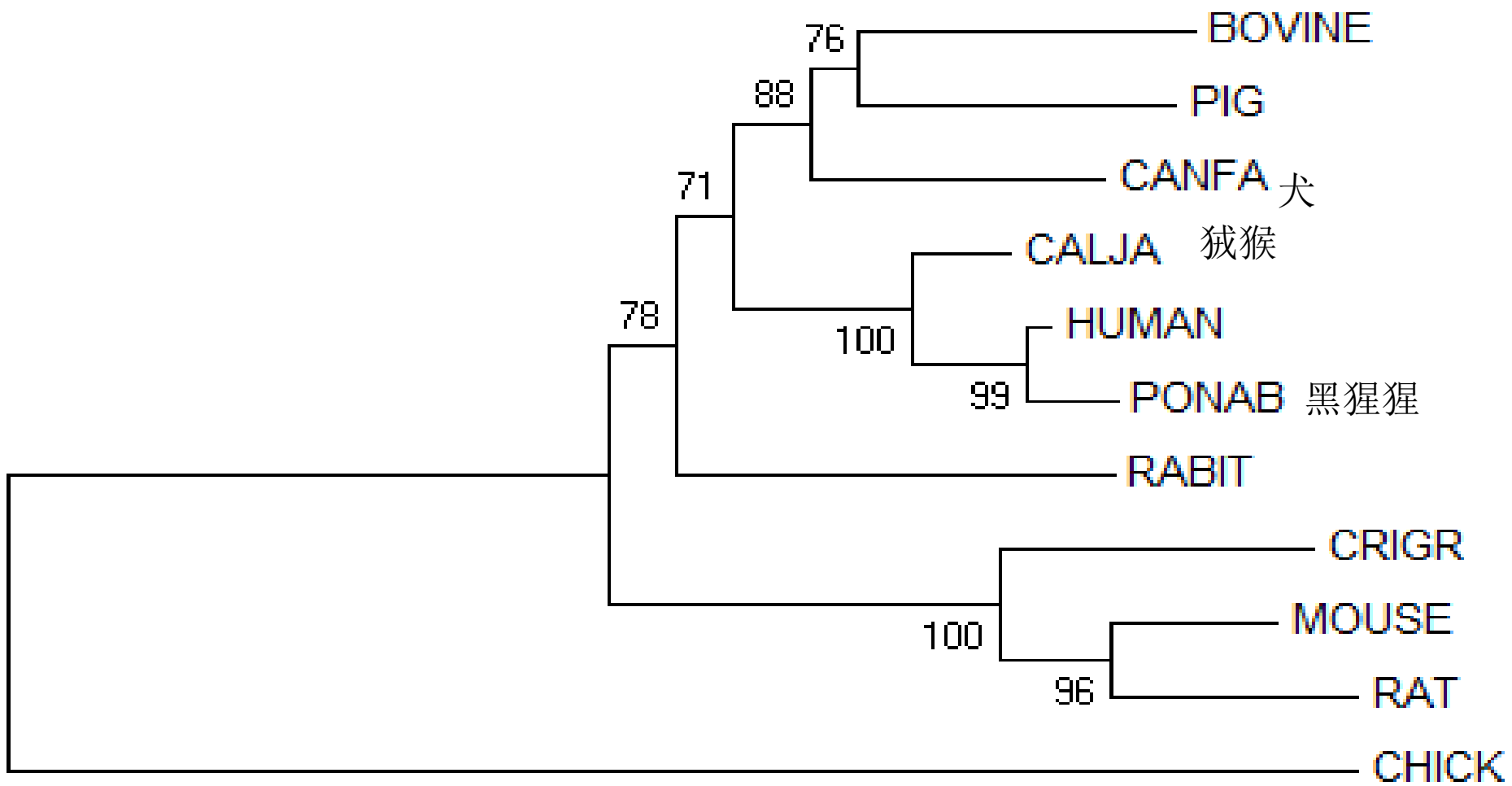
Random coil (Cc) : 265 is 29.84%

9. 用MEGA5.05软件对11个物种构建HMGCR同源序列的NJ进化树

Data Edit Search Alignment Web Sequencer Display Help

Protein Sequences

Species/Abk	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100												
1. BOVINE	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
2. CHICK	-	-	-	-	-	-	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N
3. RABIT	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
4. PIG	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
5. CALJA	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
6. HUMAN	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
7. PONAB	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
8. MOUSE	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
9. CRIGR	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
10. CANFA	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	
11. RAT	M	L	S	L	F	M	H	G	L	F	V	A	R	H	P	E	V	I	V	G	T	V	L	I	C	M	S	M	M	F	T	G	N	I	C	G	N	Y	E	C	R	F	E	E	V	L	S	D	I	I	L	I	R	T	A	I	L	Y	I	F	F	N	L	L	G	S	V	I	L	G	I	A	G	L	T	I	F	S	E	V	F	S	V	V	I	H	F	L	O	K	E	L	G	L	E	A	L	F	F	L	L	I	V	L	S	A	S	A	L	A	N	



0.01

采用needle(V6.0.1)进行两两序列比对

	IDENTITY(%)	SIMILARITY(%)
PIG	96.1	98.1
CANFA	95.9	98.0
CALJA	95.4	97.2
RABIT	94.1	97.1
HUMAN	95.4	97.0
RONAB	95.2	96.7
MOUSE	92.7	96.4
CRIGR	92.5	96.4
RAT	92.5	96.3
CHICK	84.7	91.2

10.蛋白质的三级结构预测

用**Swiss-model**将**HMGCR**与蛋白质结构数据库中的蛋白质三维结构进行匹配，以人的**HMGCR**为模板（**441-875**），输出模拟的三维结构图。



α 螺旋表示为深红色； β 折迭表示为黄色；转角（Turn）表示为淡蓝色；其它残基的颜色为白色。

InterproScan

1

888

IPR000731: Sterol-sensing 5TM box, Domain (61-218)

PS50156

IPR002202: Hydroxymethylglutaryl-coenzyme A reductase, Domain (475-871) , (464-871)

PF00368

PS50065

IPR004554: 3-hydroxy-3-methylglutaryl Coenzyme A reductase, Domain (462-871)

TIGR00533

IPR004816: Hydroxymethylglutaryl-CoA reductase (NADPH), Family (1-888)

TIGR00920

IPR009023: Hydroxymethylglutaryl-CoA reductase, NAD-binding, Domain (587-703)

SSF55035

IPR009029: Hydroxymethylglutaryl-CoA reductase, substrate-binding, Domain (441-861)

SSF56542

noIPR: unintegrated, unintegrated (46-221)

SSF82866

11.后期研究内容

拜斯亭事件后，他汀类药物长期服用的安全性备受关注，设计新型安全的**HMGCR**抑制剂十分迫切。我们根据三级结构的预测，并结合前人研究找出该酶与抑制剂的结合位点，为以后**HMGCR**抑制剂的虚拟筛选模型的建立提供依据。

参考文献

1. Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22,195-201.
2. Schwede T, Kopp J, Guex N, and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31: 3381-3385.
3. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 18: 2714-2723.
4. Sarver RW, Bills E.(2008). Thermodynamic and structure guided design of statin based inhibitors of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *J Med Chem*,51:3804-3813.
5. Park WK, Kennedy RM, and Larsen SD.(2008) Hepatoselectivity of statins: design and synthesis of 4-sulfamoyl pyrroles as HMG-CoA reductase inhibitors. *Bioorg Med Chem Lett.* 18(3):1151-1156.

致谢

- ◆首先，感谢罗老师**10**天来的辛勤付出！
- ◆其次，感谢我的组员尚利明、魏金金、张新英三位同学的相互帮助、配合！
- ◆最后，感谢研究生院给我们提供了好的学习平台！

谢谢各位!

