



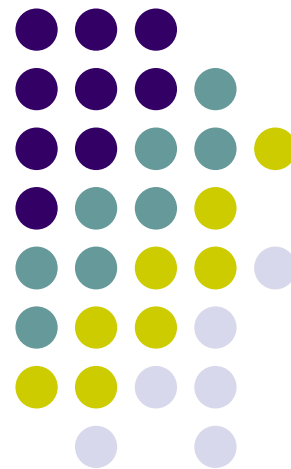
Toll样受体家族接头蛋白SARM1分析

组长、汇报人：郭小春

组员：周梦婷、张钊祥、江浩林

胡思梦、于雅馨、吴凡

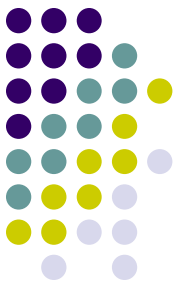
日期：2020.1.4





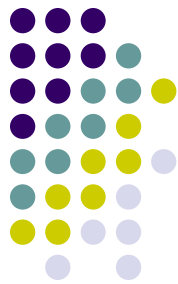
一、SARM1背景介绍

- Toll样受体 (Toll-like receptors, TLR) 是参与非特异性免疫 (天然免疫) 的一类重要蛋白质分子, 也是连接非特异性免疫和特异性免疫的桥梁。
- TLR是单个的跨膜非催化性蛋白质, 可以识别来源于微生物的具有保守结构的分子。当微生物突破机体的物理屏障, 如皮肤、粘膜等时, TLR可以识别它们并激活机体产生免疫细胞应答。



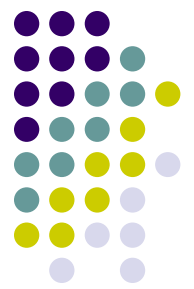
一、SARM1背景介绍

- SARM1蛋白是TLRs通路最后一个被发现的接头分子；与TLRs通路中其他四个接头分子的激活作用不同，研究表明人类SARM1专门负调控诱导IFN- β 的包含TIR结构域的接头分子依赖的TLR3和TLR4通路。
- SARM1蛋白在周围神经病变、创伤性脑损伤以及神经退行性疾病等多种神经系统疾病中发挥关键作用。研究表明，损伤可以激活SARM1并引发轴突变性，其机制扑朔迷离。在2017年国外科学家报道了SARM1的TIR结构域具有NADase活性，使得人们开始理解它的作用机制。



一、SARM1背景介绍

- 在细胞内，SARM1处于非活化状态，目前有研究报道NMN以及其结构类似物CZ-48可以使其发生构象变化，包括释放自抑制结构域和酶促结构域的二聚化等，从而使其具有酶促活性，诱导线粒体功能障碍并引起非凋亡性细胞死亡。
- 今年8月份，一篇发表在Science上的文章解析出了人类SARM1蛋白TIR结构域的三维结构，并说明了其NAD⁺催化活性。为了进一步了解人类SARM1蛋白的结构和功能，我们组开展了本次报告的分析内容。



二、SARM1基因的核酸序列分析

- 使用GENESCAN预测SARM1的核苷酸序列中可能的ORF (开放阅读框)

<http://hollywood.mit.edu/GENSCAN.html>

This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms.

This server can accept sequences up to 1 million base pairs (1 Mbp) in length. If you have trouble with the web server or if you have a large number of sequences to process, request a local copy of the program (see instructions at the bottom of this page).

Organism: Suboptimal exon cutoff (optional):

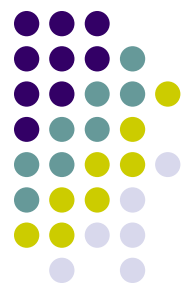
Sequence name (optional):

Print options:

Upload your DNA sequence file (upper or lower case, spaces/numbers ignored): 未选择任何文件

Or paste your DNA sequence here (upper or lower case, spaces/numbers ignored):

```
>NC_000017.11:28371694-28404049 Homo sapiens chromosome 17, GRCh38.p13
Primary Assembly
ATCTCCAGCTCAGCCGAGCCCGTGCCAGGCCACGCTTTGTTCCAGCCGCGCCTCCTCTACCCCTAAGG
CGTCGCGAGCCATCCCTCGCGCTGCTCGCTCTCTCTTTGCGCCACTCCCTGCATCTGGGCGCTGCATCAC
TTTGCAACCCGCTCCCGGATCCTGCGGACACTCCTCCCGAAACTTCTGACCGGCACCCCTGCGCTGTA
CCCTTCTTCCATTCCTCCCGCTCCATCTCTTTCCCGACCCCTCTGGGGTCCCTCTTTCCCAAAACC
CGGGTCTCTCCGCGTGGCCCGCCTCCAGGCGGGGATGTCGCCCGCGGCCCGCCATGGTCCTGAC
GCTGCTTCTTCCGCTACAAAGCTGTGTCGCTTCTTGGCCATGTCGGGCCACGGCGGGGCGGAGCGG
CTGGCGGTGCCGCGGAGATGGGGGCGGTGGCACGGGCCATGGTGGGCTGCGGGTGGCGCGGGCCCG
GCGAAGTTCGCGGGGACAGGACACCGAGGTGCAGGACGCCCTGGGAGCGCGGCTGCGGGAGCTGCAGCA
GGCCTTGTCCGCGCTGAAGCAGGCGGGGCGCGCGGGCGGTGGGCGCGGCTGGCGGAGGCTTCCAA
CTGGTGGAGGCGCTGGCTGCTGCCGCGGTGGCCCGAGGTAGCCAGGCTGTGCGACGCCATCC
```



二、SARM1基因的核酸序列分析

- 利用GENESCAN预测核酸序列中可能的ORF的结果。预测结果显示ORF起始为383bp，终止于24636bp，PolyA位点26127bp

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	Init +		383	852	470	1	2	75	105	808	0.346	76.26
1.02	Intr +		9553	10171	619	1	1	54	98	1046	0.992	94.52
1.03	Intr +		12707	12919	213	1	0	122	47	379	0.997	36.11
1.04	Intr +		13189	13280	92	0	2	85	80	183	0.981	15.99
1.05	Intr +		13390	13625	236	1	2	64	60	518	0.979	43.93
1.06	Intr +		16524	16626	103	1	1	85	94	118	0.992	11.33
1.07	Intr +		16700	16889	190	2	1	4	100	278	0.989	20.19
1.08	Intr +		24255	24376	122	2	2	110	63	168	0.999	15.79
1.09	Term +		24507	24636	130	0	1	116	41	229	0.920	18.65
1.10	PlyA +		26127	26132	6							1.05



二、SARM1基因的核酸序列分析

- 将内含子去除后的序列翻译成氨基酸序列，在NCBI蛋白质比对网站进行序列比对，发现和人类的SARM1蛋白序列是重合的。

Sequences producing significant alignments

Download Manage Columns Show 500

127 sequences selected

[GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

Sequences with E-value BETTER than threshold

select all 27 sequences selected

PSI-BLAST iteration 1

	Description	Max score	Total score	Query cover	E value	Per. Ident	Accession	Select for PSI blast	Used to build PSSM	Newly added
<input checked="" type="checkbox"/>	RecName: Full=Sterile alpha and TIR motif-containing protein 1; AltName: Full=Sterile alpha and Armadillo repeat prot	1466	1466	100%	0.0	100.00%	Q6SZW1.1	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Sterile alpha and TIR motif-containing protein 1; AltName: Full=Tir-1 homolog; Short=MyD88-5; Flags:	1345	1345	99%	0.0	94.04%	Q6PDS3.1	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Sterile alpha and TIR motif-containing protein 1; AltName: Full=Tir-1 homolog; Flags: Precursor [Sus.s	1296	1296	100%	0.0	94.48%	I3L5V6.1	<input checked="" type="checkbox"/>		
<input checked="" type="checkbox"/>	RecName: Full=Sterile alpha and TIR motif-containing protein 1; AltName: Full=Tir-1 homolog; Short=dSarm [Drosophi	600	600	84%	0.0	48.30%	Q6IDD9.1	<input checked="" type="checkbox"/>		



二、SARM1基因的核酸序列分析

- 用NetGene2（通过对特征序列“GT-AG”的分析进行直接的预测基因预测软件）进SARM1基因的内含子/外显子剪切位点识别。

<http://www.cbs.dtu.dk/services/NetGene2/>

DTU Bioinformatics
Department of Bio and Health Informatics

Services are gradually being migrated to <https://services.healthtech.dtu.dk/>.
In the near future, cbs.dtu.dk will be retired. Please try out the new site.

[Home](#)

NetGene2 Server

The NetGene2 server is a service producing neural network predictions of splice sites in human, *C. elegans* and *A. thaliana* DNA.

Instructions	Output format	Abstract
--------------	---------------	----------

SUBMISSION

Submission of a local file with a single sequence:

File in **FASTA** format 未选择任何文件

Human
 C. elegans
 A. thaliana

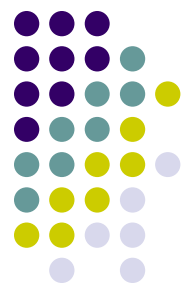
Submission by pasting a single sequence:

Sequence name

Human
 C. elegans
 A. thaliana

Sequence

```
AGGATTGTTTGGAGCCAGGAGTTTTCAGCCAGCCTGGGCAATATAATGGGATCCTGCTCTACAAAAGTAA
AAGAAAACAACAAAACAAAATAATCAGCCAGACATGGTGTACATGCCCTAGTCCACGCTACTCAGGTGGC
TGAGTTGGGAGGATGAATTTAACTAGGTGGCCAGGCTGCAGTGAAGCTGTGAGCTTATGGCATGTTCCG
GACTCAGGTAGGTTCTGCAGTGGACACAAAATCACATGCCACTATGGTCACTGTGGTTCGCCCTTTGGGCAAG
GACTCCAATTGCTGGCTAGGCCCTTACCTCTTTGGCAGTAATGTCACCTGTGGAGGACATCACTTTTTTTT
```

二、SARM1基因的核酸序列分析

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
309	0	+		0.34	CGCCTCCAG	GCCGGGGATG		
372	0	+		0.07	CGCCTACAAG	CTGTGTCGCT		
496	1	+		0.19	CCCCGCGAAG	TGTCGCCGGG		
511	1	+		0.43	CCGGGGGCAG	GCACCGAGGT		
519	0	+		0.34	AGGCACCGAG	GTGCAGGACG		
525	0	+		0.34	CGAGGTGCAG	GACGCCCTGG		
537	0	+		0.19	CGCCCTGGAG	CGCGCGCTGC		
940	2	+		0.25	TCTCTGTTAG	GGGTCAGCCT		
963	1	+		0.43	TGTTTCACAG	ATAAGGAAAC		
1036	2	+		0.33	TTGCTCCCAG	GGCTAGTGGG		
1286	1	+		0.56	TCCTCCCCAG	GCTGTAAGGC		
1378	0	+		0.19	CTTGCAACAG	GCCAGACATG		
1383	2	+		0.17	AACAGGCCAG	ACATGATCTC		
1638	1	+		0.38	CTGATGCCAG	AGCAGACCAA		
2019	0	+		0.26	TGCCCTGCAG	GCTTGCTATG		
3361	0	+		0.80	TTCTTCTCAG	GATGCTATTG		
3509	2	+		0.29	ATATTACAG	TCCACCCCTC		
4218	1	+		0.49	ATCCTTCCAG	AGAGAGTCTC		
4553	1	+		0.27	TGGGCCTCAG	GTAAGTCCTC		
5439	1	+		0.25	TACATTCTAG	GCATTTATGC		
6025	0	+		0.33	TTTTTTTTAG	ACGGAGTCTC		
6282	1	+		0.07	GGGATTACAG	GCATGAGCCA		

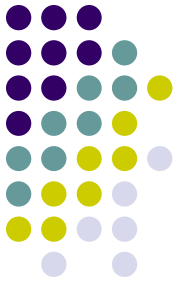
The sequence: NC_000017.11_28371694-28404049 has the following composition:

Length: 32356 nucleotides.
24.9% A, 24.5% C, 25.9% G, 24.8% T, 0.0% X, 50.4% G+C

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
810	2	+		1.00	CTGAGAACC	GTGAGCGCGC		H
2003	1	+		0.37	GTC AAGGCTG	GTAGGTGTC		
2064	2	+		0.41	TCCTGGAGA	GTGAGTACAG		
2247	0	+		0.54	AGAGCACAAC	GTGCGTGGTT		
3053	2	+		0.37	GGAAGCTGAG	GTGGTGGAT		
3470	1	+		0.30	ACAACACCCA	GTGAGTCCC		
3973	1	+		0.34	CACACAAAAG	GTATACAGTT		
4397	1	+		0.83	TGGGAGCTGG	GTGAGTAGTT		
4463	0	+		0.41	ATAGGGAAAA	GTGAGGGCTG		
4554	2	+		0.49	TGGGCCTCAG	GTAACTCCTC		
4900	1	+		0.46	AAAAGAAAAA	GTAAAGGTGT		
5621	1	+		0.56	ACTGGGCCAG	GTGAGGTGGT		
5888	0	+		0.47	CCGGACTCAG	GTAGGTCTG		
6331	1	+		0.51	ATTGTGGAG	GTGAGTAAAG		
7493	0	+		0.53	TTTATTCAAG	GTATGTGTGT		
7586	1	+		0.34	CTAGGAAAGG	GTAAATTTCT		
8032	2	+		0.37	ATAAAATAAG	GTAAATGAGTT		
9112	0	+		0.34	TATGAGGCAG	GTATTCTAT		
9150	2	+		0.37	GGAAAGCAAG	GTAGAGAGAT		
9411	1	+		0.41	TATGAGGCCG	GTGAGGCCCA		
9427	2	+		0.41	CCCAGAGAGG	GTAGTGACC		

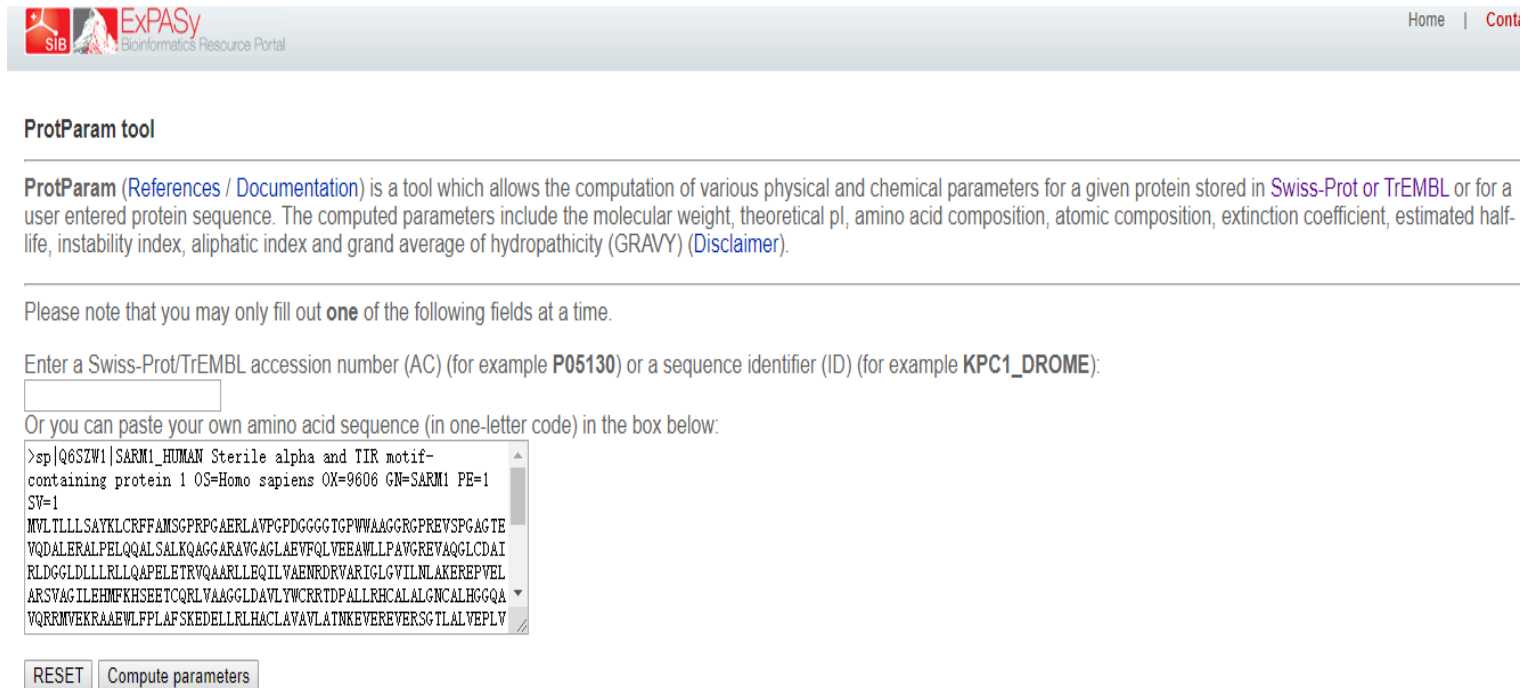
通过核酸序列分析，对SARM1蛋白基因的结构和剪接表达有了一定的认识。



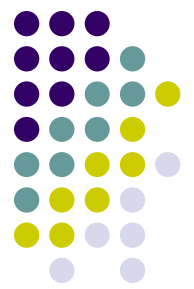
三、SARM1蛋白结构分析

- 利用ProtParam预测小鼠Ythdf2蛋白基本理化性质。

<https://web.expasy.org/protparam/>



The screenshot shows the ProtParam tool interface. At the top, there is a header with the SIB logo and 'ExpASy Bioinformatics Resource Portal'. Below the header, the title 'ProtParam tool' is displayed. The main text describes the tool's function: 'ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer)'. Below this, a note states: 'Please note that you may only fill out one of the following fields at a time.' There are two input options: 'Enter a Swiss-Prot/TrEMBL accession number (AC) (for example P05130) or a sequence identifier (ID) (for example KPC1_DROME):' followed by an empty text box, and 'Or you can paste your own amino acid sequence (in one-letter code) in the box below:'. The second option is selected, and a text area contains the following sequence: '>sp|Q6SZW1|SARM1_HUMAN Sterile alpha and TIR motif-containing protein 1 OS=Homo sapiens OX=9606 GN=SARM1 PE=1 SV=1 MVLTLILL SAYKLCRFFAMSGPRPGAERLAVPGPDGGGGTGFVWAAGGRGPREVSPGAGTE VQDALERALPELQQALSALKQAGGARAVGAGLAEVFQLVEEAWLLPAVGREVAQGLCDAI RLDGGGLDLLRLQAPELETRVQAARLLLEQILVAENRDRVARIGLVILNAKEREPVEL ARSVAGILEHMFKHSEETCQRLVAAGGLDAVLYWCRRTDPALLRHCALALGNCALHGGQA VQRRMVEKRAAEVLFPLAFSKEDELLRLHACLAVALATNKEVEREVERSGTLALVEPLV'. At the bottom, there are two buttons: 'RESET' and 'Compute parameters'.



SARM1蛋白基本理化性质

- 利用ProtParam预测结果显示，该蛋白质的氨基酸数量有724个，相对分子质量为79388.23，理论的pH值为6.14，正电荷残基数为91，负电荷残基数为84。

Number of amino acids: 724

Molecular weight: 79388.23

Theoretical pI: 6.14

Amino acid composition: [CSV format](#)

Ala (A)	72	9.9%
Arg (R)	55	7.6%
Asn (N)	12	1.7%
Asp (D)	36	5.0%
Cys (C)	18	2.5%
Gln (Q)	34	4.7%
Glu (E)	55	7.6%
Gly (G)	62	8.6%
His (H)	14	1.9%
Ile (I)	25	3.5%
Leu (L)	103	14.2%
Lys (K)	29	4.0%
Met (M)	10	1.4%
Phe (F)	23	3.2%
Pro (P)	33	4.6%
Ser (S)	46	6.4%
Thr (T)	28	3.9%
Trp (W)	11	1.5%
Tyr (Y)	10	1.4%
Val (V)	48	6.6%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 91
 Total number of positively charged residues (Arg + Lys): 84

Atomic composition:

Carbon	C	3510
Hydrogen	H	5647
Nitrogen	N	1003
Oxygen	O	1037
Sulfur	S	28

Formula: C₃₅₁₀H₅₆₄₇N₁₀₀₃O₁₀₃₇S₂₈

Total number of atoms: 11225



SARM1蛋白基本理化性质

- 预测该蛋白的分子式 $C_{3510}H_{5647}N_{1003}O_{1037}S_{28}$ ，脂肪系数为98.12，亲水性平均值为-0.108。

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.

Ext. coefficient 76525
Abs 0.1% (=1 g/l) 0.964, assuming all pairs of Cys residues form cystines

Ext. coefficient 75400
Abs 0.1% (=1 g/l) 0.950, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

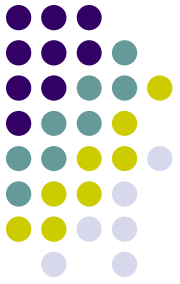
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 42.23
This classifies the protein as unstable.

Aliphatic index: 98.12

Grand average of hydropathicity (GRAVY): -0.108



预测SARM1蛋白跨膜区

(1) 利用TMHMM预测蛋白跨膜区域

<http://www.cbs.dtu.dk/services/TMHMM-2.0/>

DTU Bioinformatics
Department of Bio and Health Informatics

Services are gradually being migrated to <https://services.healthtech.dtu.dk/>.
In the near future, cbs.dtu.dk will be retired. Please try out the new site.

[Home](#)

TMHMM Server v. 2.0

Prediction of transmembrane helices in proteins

[Instructions](#)

SUBMISSION

Submission of a local file in **FASTA** format (HTML 3.0 or higher)

未选择任何文件

OR by pasting sequence(s) in **FASTA** format:

```
>sp|Q6SZW1|SARM1_HUMAN Sterile alpha and TIR motif-containing protein
1 OS=Homo sapiens OX=9606 GN=SARM1 PE=1 SV=1
MVL TLLLSAYKLCRFFAMSGPRPGAERLAVPGDGGGGTGPWWAAGGRGPREVSPGAGTE
VQDALERALPELQQALSALKQAGGARAVGAGLAEVYFQLVEEAWLLPAVGREVAQGLCDAI
RLDGGDLLLRLQLAPELETQVQAARLLBQILVAENRDRVARIGLVILMLAKEREPEVEL
```

Output format:

- Extensive, with graphics
- Extensive, no graphics
- One line per protein

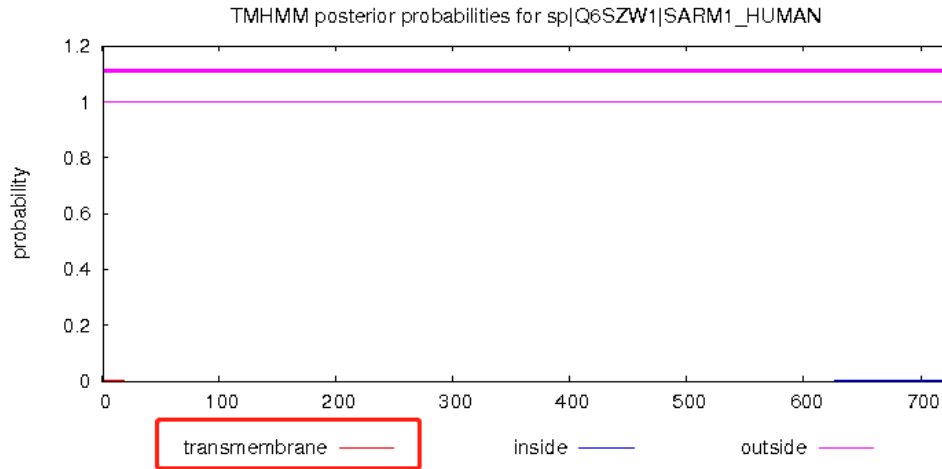


预测SARM1蛋白跨膜区

TMHMM result

[HELP](#) with output formats

```
# sp|Q6SZW1|SARM1_HUMAN Length: 724
# sp|Q6SZW1|SARM1_HUMAN Number of predicted TMHs: 0
# sp|Q6SZW1|SARM1_HUMAN Exp number of AAs in TMHs: 0.00694
# sp|Q6SZW1|SARM1_HUMAN Exp number, first 60 AAs: 0.00338
# sp|Q6SZW1|SARM1_HUMAN Total prob of N-in: 0.00026
sp|Q6SZW1|SARM1_HUMAN TMHMM2.0 outside 1 724
```



[plot](#) in postscript, [script](#) for making the plot in gnuplot, [data](#) for plot

- 利用TMHMM预测SARM1蛋白跨膜结构的结果是该蛋白不存在跨膜结构



预测SARM1蛋白跨膜区

(2) 利用HMMTOP预测蛋白跨膜结构

<http://www.enzim.hu/hmmtop/html/submit.html>

Please try our new topology prediction method, called **CCTOP!**

Home | Documentation | Help | Download | Advanced | Submit

Copyright © G. E. Tusnady, 2001

Please paste the sequence to the text area below.

```
>sp|Q0SZW1|SARM1_HUMAN Sterile alpha and TIR motif-
containing protein 1 OS=Homo sapiens OX=9606 GN=SARM1 PE=1
SW=1
MVL TLLLSAYKLCRFFAMSGPRPGAERLAVPGPDGGGGTGPWWAAGGRGPREVSPGAGTE
VQDALERALPELQQALSALKQAGGARAVGAGLAEVFQLVEEAWLLPAVGREVAQGLCDAI
```

Submit Clear



预测SARM1蛋白跨膜区

Protein: noname
 Length: 794
 N-terminus: OUT
 Number of transmembrane helices: 1
 Transmembrane helices: 70-87

Total entropy of the model: 17.0154
 Entropy of the best path: 17.0164

The best path:

```

seq SPQSWSARMH MANSTERILE ALPHAANDTI RMTIFCNTAI NINGPRTEIN 50
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq SHMSAPIENS GNSARMPESV MVLTLLLSAY KLKCRFFAMSG PRPGAERLAV 100
pred 0000000000 0000000000H HHHHHHHHHH HHHHHHHHiii iiiiiiiiii

seq PGPDGGGGTG PWWAAGGRGP REVSPGAGTE VQDALERALP ELQQALSALK 150
pred iiIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq QAGGARAVGA GLAEVFQLVE EAWLLPAVGR EVAQGLCDAI RLDGGLDLLL 200
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq RLLQAPELET RVQAARLLEQ ILVAENRDRV ARIGLGVILN LAKEREPVEL 250
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq ARSVAGILEH MFKHSEETCQ RLVAAGGLDA VLYWCRRTDP ALLRHICALAL 300
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq GNCALHGGQA VQRRMVEKRA AEWLFPLAFS KEDELLRLHA CLAVAVLATN 350
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq KEWEREVERS GTLALVEPLV ASLDPGRFAR CLVDASDTSQ GRGPDDLQRL 400
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq VPLLDNSRLE AQCIGAFYLC AEAAIKSLQG KTKVFSDIGA IQSLKRLVSY 450
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq STNGTKSALA KRALRLGEE VPRPILPSVP SWKEAEVQTW LQQIGFSKYC 500
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII
  
```

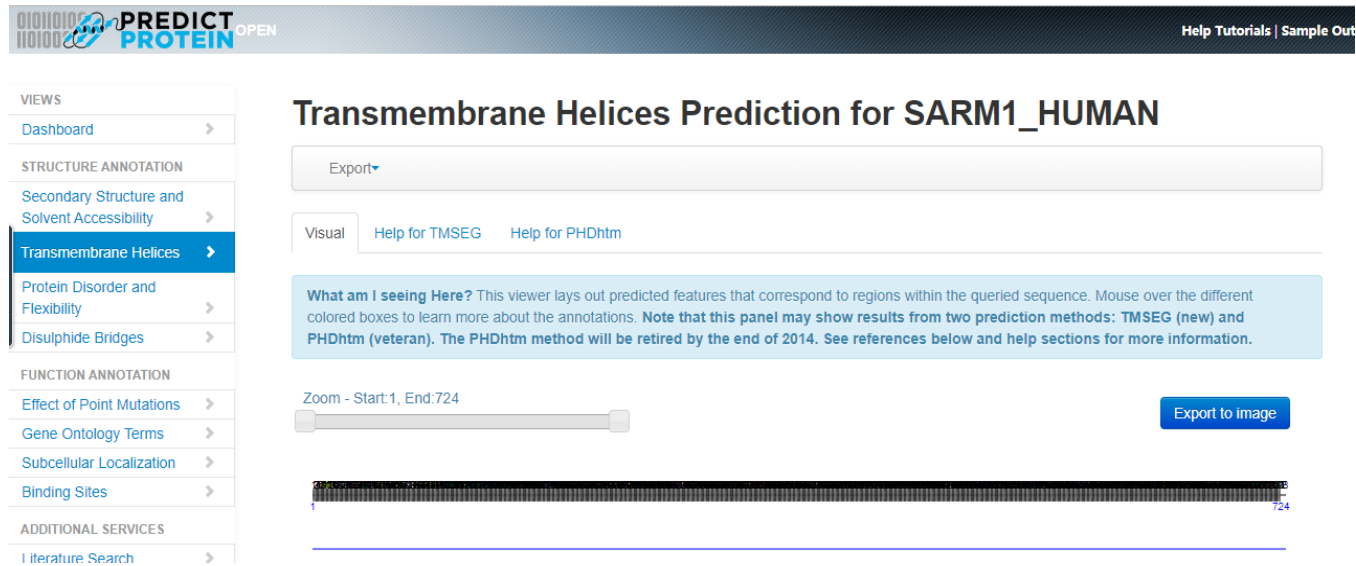
- 可以发现用HMMTOP预测蛋白跨膜结构的结果是，该蛋白在70-87号氨基酸可能存在1个跨膜螺旋。



预测SARM1蛋白跨膜区

(3) 利用PredictProtein预测蛋白质跨膜结构

<http://www.predictprotein.org/>



PREDICT PROTEIN OPEN Help Tutorials | Sample Out

Transmembrane Helices Prediction for SARM1_HUMAN

Export

Visual Help for TMSEG Help for PHDhtm

What am I seeing Here? This viewer lays out predicted features that correspond to regions within the queried sequence. Mouse over the different colored boxes to learn more about the annotations. **Note that this panel may show results from two prediction methods: TMSEG (new) and PHDhtm (veteran). The PHDhtm method will be retired by the end of 2014. See references below and help sections for more information.**

Zoom - Start:1, End:724

Export to image

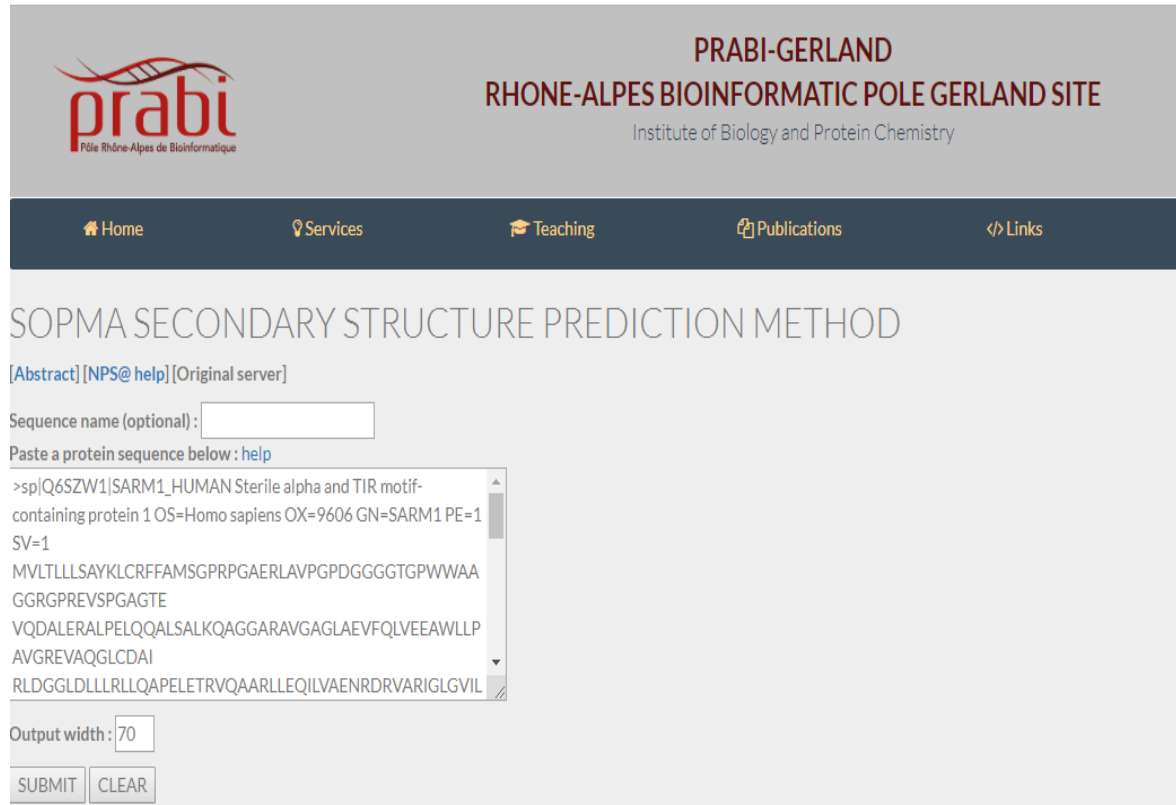
- 利用PredictProtein预测蛋白跨膜结构的结果。预测SARM1蛋白不存在跨膜结构。
- 综上，认为SARM1蛋白不存在跨膜结构。



预测SARM1蛋白的二级结构

(1) 利用SOPMA预测蛋白质二级结构

https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html



PRABI-GERLAND
RHONE-ALPES BIOINFORMATIC POLE GERLAND SITE
Institute of Biology and Protein Chemistry

Home Services Teaching Publications Links

SOPMA SECONDARY STRUCTURE PREDICTION METHOD

[Abstract] [NPS@ help] [Original server]

Sequence name (optional):

Paste a protein sequence below : [help](#)

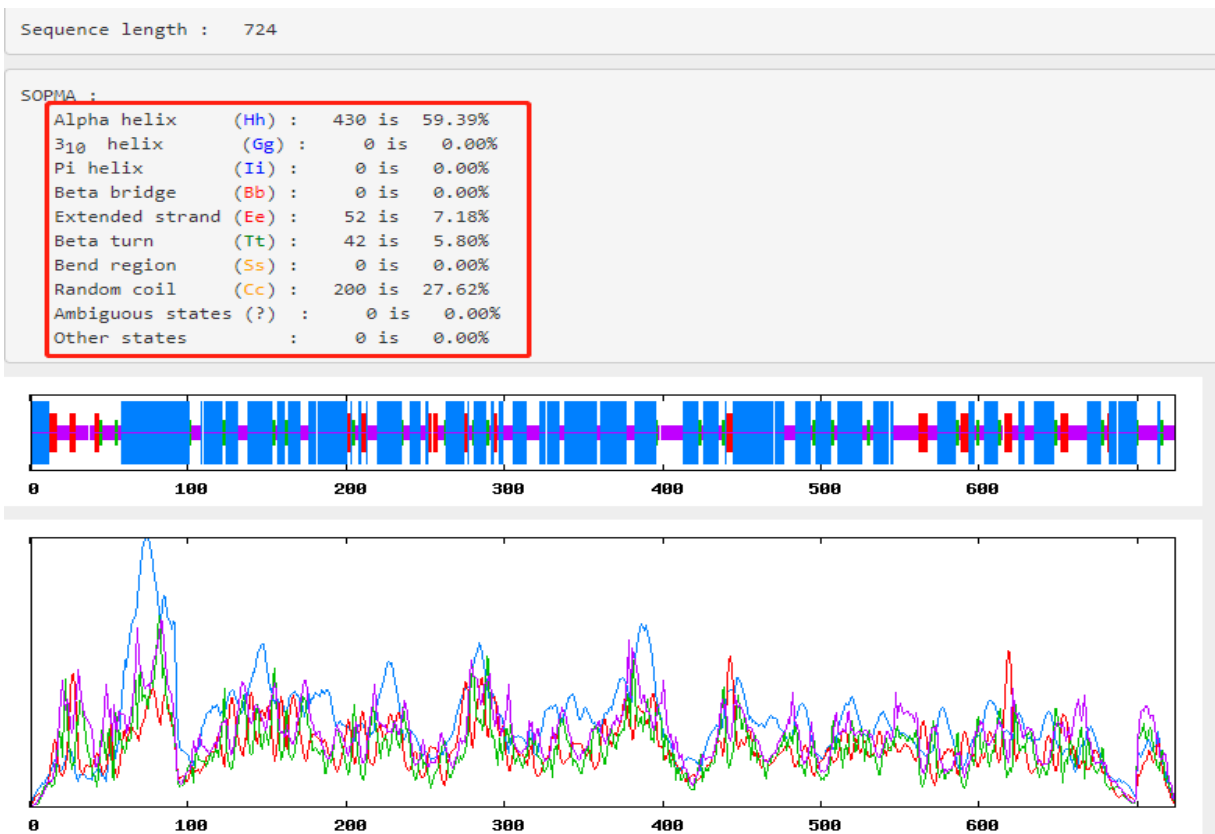
```
>sp|Q6SZW1|SARM1_HUMAN Sterile alpha and TIR motif-  
containing protein 1 OS=Homo sapiens OX=9606 GN=SARM1 PE=1  
SV=1  
MVLTLSSAYKLCRFFAMSGPRPGAERLAVPGPDGGGGTGPWWAA  
GGRGPREVSPGAGTE  
VQDALERALPELQQALSALKQAGGARAVGAGLAEVFQVVEEAWLLP  
AVGREVAQGLCDAI  
RLDGGDLLLRLQLAPELETRVQARLLEQILVAENRDRVARIGLVIL
```

Output width:

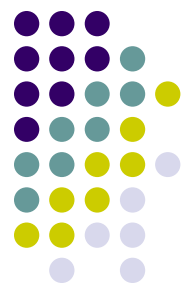
SUBMIT CLEAR



预测SARM1蛋白的二级结构



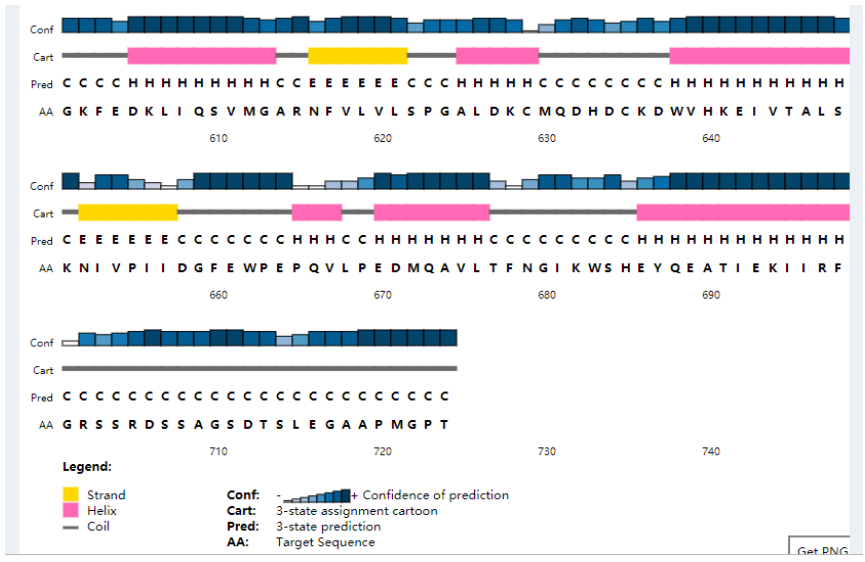
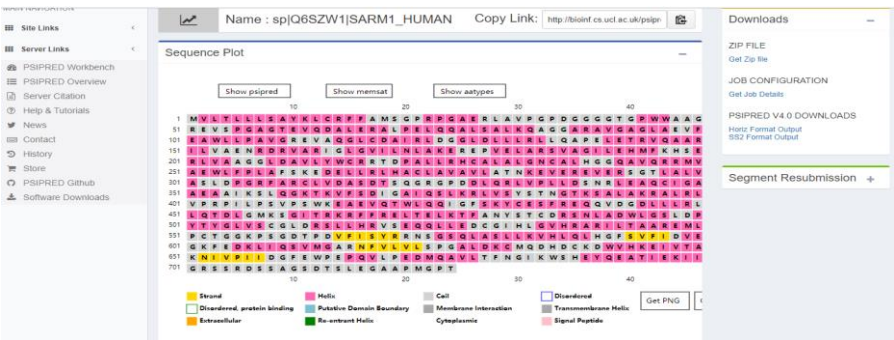
- 从上图预测SARM1蛋白的二级结构以 α 螺旋为主，其次为无规则卷曲， β 折叠和 β 转角较少，卷曲、 3_{10} 螺旋、 π 螺旋、 β 桥则没有。



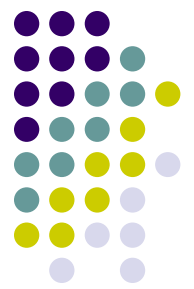
预测SARM1蛋白的二级结构

(2) 利用PSIpred预测蛋白质二级结构

<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>



- 左：全面的二级结构分布。黄色表示折叠，红色表示螺旋，表示预测出有42个螺旋和4个折叠。
- 右：具体的二级结构分布，预测该蛋白螺旋最多（红色柱状），卷曲数量其次（黑色横线），折叠最少（黄色柱状）。



预测SARM1蛋白的二级结构

(3) PredictProtein预测蛋白质二级结构

<http://www.predictprotein.org/>

```
>sp|Q6SZW1|SARM1_HUMAN Sterile alpha and TIR motif-containing protein 1 OS=Homo sapiens OX=9606 GN=SARM1 PE=1 SV=1
MVLTLLLSAYKLCRFFAMSGPRPGAERLAVPGPDGGGOTGPWAAAGRGRPREVSPGAGTE
VDALERALELQDALSAIKGAGARAVGAGLVEVQLVEEAWLFLVAVGREGVQGLCDAI
RLDGGLLDLLRLQAPLETRVQAAARLEQLVAENRDRVARIGLVILNAKEREPVEL
ARSVAGILEHMFKHSEETCQRLVAAGGLDAVLYWCRRTDPALLRHCALALGNALHGGQA
VORRMVEKRAAEWLFPFLAFSKEDELLRLHACLAVAVLATNKEVEREVERSGTLALVEPLV
ASLDPGRFARCLVDSQTSQGRGPDQLGRVPLLDNRLAQCIGAPYLCAEAAIKSLQG
KTRVFSQGAQSLKRLVSYSTNGTKSALAKRALRLLGEEVFRPLPSVPSWAEVQVW
LQIQGFSKYCESFREQQVDGDLRLTEEELQTLGMKSGITRKRFRRELTKTFANYS
TCDRSNLADWLSLDRFRQYTYGLVSCGLDRSLHRVSEQQLLEDCGIHLGVHRARILT
AAREMLHSPLPCTGGKPSGDTDPVFSYRRNSQSALSKVHLQLHGFVDFVDEKLEA
GKFEDKLIQSVMGARNFVVLSPGALDKCGMDHCKDWHVKEIVTALSCGNVPIIDGF
EHWPEQVLPEDMQAVLTNFKWHSHEYQEATIKIRFLQGRSSRDSAGSDTSLGGAAP
MGPT
```

Secondary Structure and Solvent Accessibility Prediction for SARM1_HUMAN

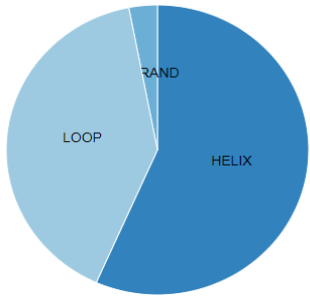


- 第一二行里红色块表示螺旋，蓝色块表示折叠。
- 第三行里蓝色块为暴露，黄色块为埋藏、空白处（除最右边一块外）为中间状态。

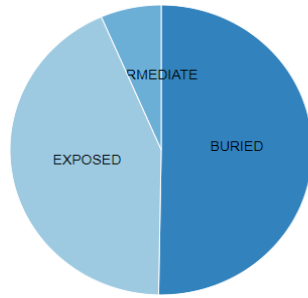


预测SARM1蛋白的二级结构

Secondary Structure Composition



Solvent Accessibility



Proteins can be classified as **mixed** given the following classes:

- 'all-alpha': %H > 45% AND %E < 5%
- 'all-beta': %H < 5% AND %E > 45%
- 'alpha-beta': %H > 30% AND %E > 20%
- 'mixed': All others

- 二级结构组成的饼状图，二级结构中以螺旋和环区为主。
- 溶剂可及性分析的饼状图，各蛋白质残基溶剂的可及性埋藏区多于暴露区。
- 二硫键预测，该蛋白质没有二硫键。

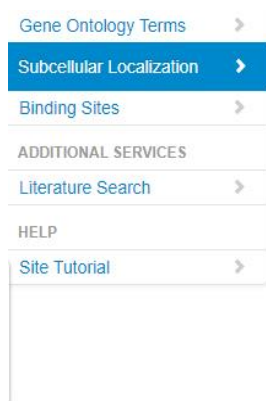
The screenshot shows the PREDICT PROTEIN web interface. The main heading is "Disulphide Bridges Prediction for SARM1_HUMAN". There is an "Export" button and a "Visual" tab. A text box explains: "What am I seeing Here? This viewer lays out predicted features that correspond to regions within the queried sequence. Mouse over the different colored boxes to learn more about the annotations". Below this is a zoom slider set to "Start 1, End 724" and an "Export to image" button. At the bottom, a protein sequence bar is visible with a red highlight at the end.



预测SARM1蛋白的亚细胞定位

(1) PredictProtein预测SARM1蛋白的亚细胞定位

<http://www.predictprotein.org/>



Predicted localization for the Eukarya domain: Cytoplasm (GO term ID: [GO:0005737](#)) Prediction confidence 27

- 该蛋白定位在细胞质内。



预测SARM1蛋白的亚细胞定位

(2) 利用Euk-mPLOC进行SARM1蛋白亚细胞定位

<http://www.csbio.sjtu.edu.cn/bioinf/euk-multi/#>

Euk-mPLOC: Predicting subcellular localization of eukaryotic proteins including those with multiple sites

[Read Me](#) | [Data](#) | [Citation](#) | [Download](#) |

Your input sequence(724aa) is:

```
>PSARM1
MVLTLILLSAYKLCRFFAMSGPRPGAERLAVPGPDGGGGTGPWWAAGGRGPREVSPGAGTE
VQDALERALPELQQALSALKQAGGARAVGAGLAEVFQLVEEAWLLPAVGREVAQGLCDIAI
RLDGGGLDLLRLQLAPELETRVQAARLLEQILVAENRRDRVARIGLVILNLAKEREPVEL
ARSVAGILEHMFKHSEETCQRLVAAGGLDAVLYWCRRTDPALLRHCALALGNCALHGGQA
VQRRMVEKRAAEWLFPLAFSKEDELLRLHACLAVAVLATNKEVEREVERSGTLALVEPLV
ASLDPGRFARCLVDASDTSQGRGPDDLQRLVPLLDNRLEAQICIGAFYLCAEAAIKSLQG
KTKVFSDIGAIQSLKRLVSYSTNGTKSALAKRALRLLGEEVPRPILPSVPSWKEAEVQTW
LQQIGFSKYCESFREQQVDGDLRLLEELQTDLGMKSGITRKRFFRELTCLKTFANYS
TCDRSNLADWLGSLDPRFRQYTYGLVSCGLDRSLLHRVSEQLLEDCEGIHLGVHRARILT
AAREMLHSPLPCTGGKPSGDTDPDVFISYRRNSGSQLASLLKVHLQLHGFVSFIDVEKLEA
GKFEDKLIQSVMGARNFVLVLSPGALDKCMQDHDCKDWVHKEIVTALSCKGNIVPIIDGF
EWPEPQVLPEDMQAVLTFNGIKWSHEYQEATIEKIIRFLQGRSSRDSSAGSDTSLEGAAP
MGPT
```

----- Euk-mPLOC Computation Result -----

Accession number	Predicted location(s)	Prediction approach
PSARM1	яя Cytoplasm	By fusing PseAA composition

[Home Page](#)

- 该蛋白定位在细胞质内。



预测SARM1蛋白的亚细胞定位

- 通过Uniport数据库查找人类的SARM1蛋白：

neXtProt¹ NX_Q6SZW1

Display

Entry

Publications

Feature viewer

Feature table

None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains

Subcellular location¹

Mitochondrion

UniProt annotation

GO - Cellular component

Mitochondrion

Mitochondrion ⓘ 2 Publications ▾

Other locations

Cytoplasm ⓘ 1 Publication ▾

axon ⓘ By similarity ▾

dendrite ⓘ By similarity ▾

synapse ⓘ By similarity ▾

Note: Associated with microtubules. ⓘ By similarity ▾

Manual annotation Automatic computational assertion

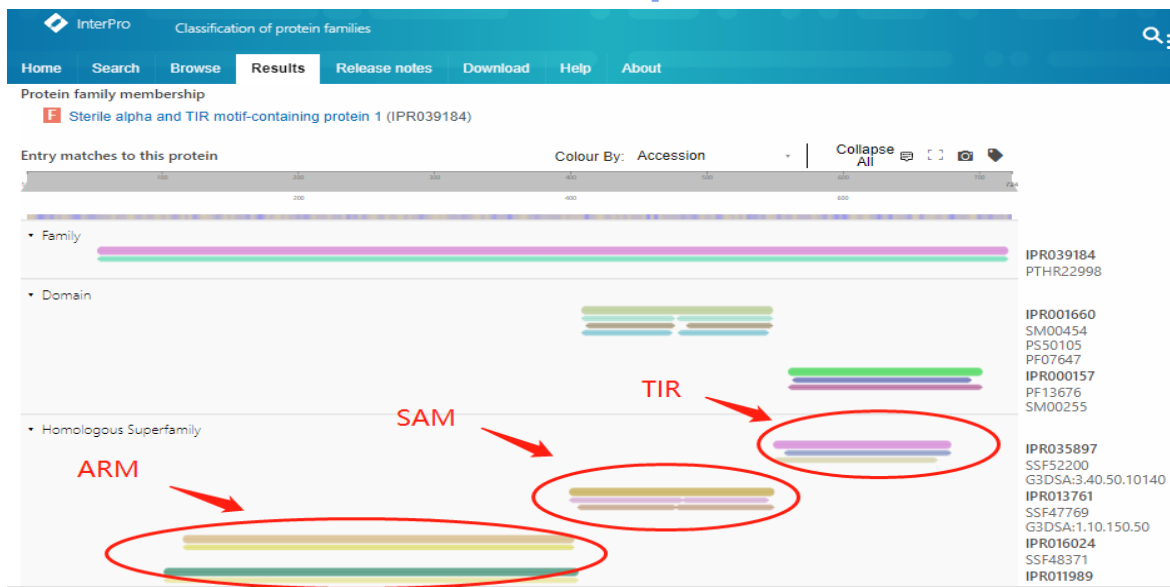
- 显示其定位在线粒体、细胞质中，还在轴突、树突中存在。



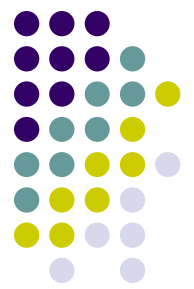
分析SARM1蛋白结构域

(1) InterProScan工具分析SARM1蛋白结构域

<http://www.ebi.ac.uk/interpro/>



- 预测该蛋白在在116-402号氨基酸处是ARM 结构域,403-549处是SAM结构域, 550-679是TIR结构域。



分析SARM1蛋白结构域

(2) 利用CDD和SMART分析SARM1蛋白结构域

- CDD: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=cdd>



NCBI Conserved Domains

Search for Conserved Domains within a protein or codina nucleotide sequence

Enter **protein** or **nucleotide** query as accession, gi, or sequence in FASTA format. For multiple protein queries, use [Batch CD-Search](#).

>sp|Q6SZW1|SARM1_HUMAN Sterile alpha and TIR motif-containing protein OS=Homo sapiens OX=9606 GN=SARM1 PE=1 SV=1

MVLTLLLSAYKLCRFFAMSGPRPGEARLAVGFDGGGGTGFVWAAAGGRGPREVSPGAGTEVDDALERALPELQALSAKQAGGARAVAGAGLAEVFGLVEEAWLLPAVGEVAQGLCDARLDGGLDLLRLRLQAPPELETRVQAARLLEQLVAENRDRVARIGLGVINLAKEREPEVELARSVAGILEHMFKHESETQRLVAAGGLDAVLVWCRDTPALLRHCAALGNLALHGGQAVQRRMVEKRAAEWLFFLAFSKEDELLRLHACLAVAVLATNKEVEREVERSGTLALVEPLVASLDPGRFARCLVDASDTSQGRGPDQLRLVFLDLSNRLEAQCIQAFYLCAEAAIKSLQGKTRVPSDIDGATQSLKRLVSYSTNGTKSALAKRALLLGEEVFRPILFVSPVSKAEAEVQTVLQQIGFSKYCESFREQQVGGDLLRLTEELQTDLGMKSGITRRKFFRELTAKTFANYS TCDRSNLADWLGSGLDPRFRQITTYGLVSCGLDRSLLRHVSEQLLEDCCGHLGVHARILTAAREMLHSLFPCGKPGSDTDFVFTSYRMSGSLASLKVHLQLHGFVFTVDEKLEA GKFEDKILQVMGARGNFFVLVSGALDKQMGDDCKDCKWTRKELVIALSCGMIVETIDDFENPEPQVLPEDMQAVLTFNGIKMSHEYQATIEKTIIRFLQQRSSRDSSAGSDTSLEGAARMGPT

Conserved domains on [sp|Q6SZW1]

SARM1_HUMAN Sterile alpha and TIR motif-containing protein OS=Homo sapiens OX=9606 GN=SARM1 PE=1 SV=1

Protein Classification

SAM_SARM1-like_repeat1 and SAM_SARM1-like_repeat2 domain-containing protein (domain architecture ID 10175649)
protein containing domains SAM_SARM1-like_repeat1, SAM_SARM1-like_repeat2, and TIR

Graphical summary Zoom to residue level [show extra options](#)

Query seq. Specific hits Superfamilies

SAM_SARM1-like SAM_SARM1-like TIR
SAM_superfam1 SAM_superfam1 TIR_2 superfamily

List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	SAM_SARM1-like_repeat2	cd09502	SAM domain of SARM1-like, repeat 2; SAM (sterile alpha motif) domain repeat 2 of SARM1-like ...	478-547	7.48e-37
[+]	SAM_SARM1-like_repeat1	cd09501	SAM domain of SARM1-like proteins, repeat 1; SAM (sterile alpha motif) domain repeat 1 of ...	409-477	4.82e-30
[+]	TIR	smart00255	Toll - interleukin 1 - resistance;	561-702	4.35e-22



分析SARM1蛋白结构域

- SMART: <http://smart.embl-heidelberg.de/>

SMART MODE: NORMAL GENOMIC

Domains within *Homo sapiens* protein SARM1_HUMAN (Q6SZW1)

Sterile alpha and TIR motif-containing protein 1

Length 724 aa

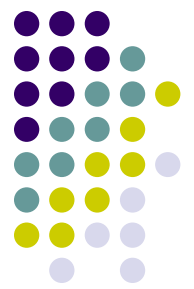
Source database UniProt

Identifiers SARM1_HUMAN, Q6SZW1, ENSP00000468032.2, ENSP00000468032, O60277, Q7LGG3, Q9NXY5, K7AVX3_PANTR, K7AVX3, H2QCH4, G3R687_GORGO, G3R687, A0A2J8TMA8_PONAB, A0A2J8TMA8, A0A2J8LI04_PANTR, A0A2J8LI04, Q0D2N8_HUMAN, Q0D2N8

Source gene ENSG00000004139

Alternative splicing SARM1_HUMAN, J3QRE0_HUMAN, J3KSG7_HUMAN, J3KRZ6_HUMAN

- 预测结果显示该蛋白在409-477和478-547号氨基酸存在两个SAM结构，在561-702存在TIR结构域。



SARM1蛋白疏水性分析

利用ProtScale进行疏水性分析

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

SEQUENCE LENGTH: 724

Using the scale **Hphob. / Kyte & Doolittle**, the individual values for the 20 amino acids are:

Ala: 1.800 Arg: -4.500 Asn: -3.500 Asp: -3.500 Cys: 2.500 Gln: -3.500
 Glu: -3.500 Gly: -0.400 His: -3.200 Ile: 4.500 Leu: 3.800 Lys: -3.900
 Met: 1.900 Phe: 2.800 Pro: -1.600 Ser: -0.800 Thr: -0.700 Trp: -0.900
 Tyr: -1.300 Val: 4.200 : -3.500 : -3.500 : -0.490

Weights for window positions 1,...,9, using **linear weight variation model**:

1	2	3	4	5	6	7	8	9
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
edge			center					edge



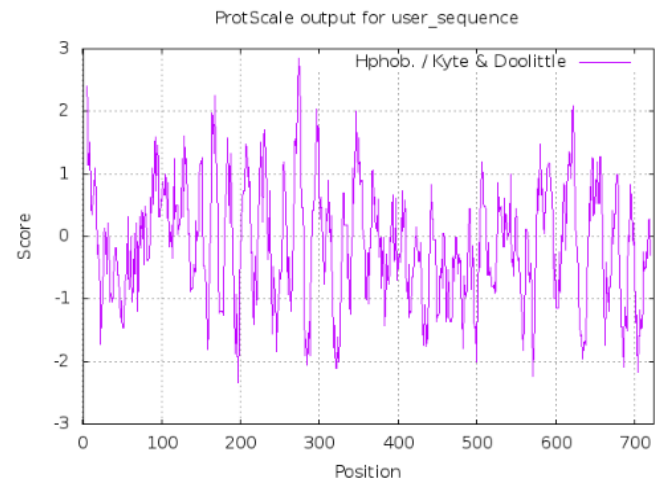
ProtScale
 ProtScale [Reference / Documentation] allows you to compute and represent the profile produced by any amino acid scale on a selected protein.

An **amino acid scale** is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity s secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids provides 57 predefined scales entered from the literature.

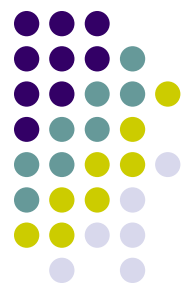
Enter a UniProtKB/Swiss-Prot or UniProtKB/TrEMBL accession number (AC) (e.g. **P05130**) or a sequence identifier (ID) (e.g. **KPC1_DROME**):

Or you can paste your own sequence in the box below:

```
LCVHRAEILT
AAEEMLHSPFCGGKPSGDTDFWFISTYRNSGSQLASLLKVLQLHGFSS
VFIDVEKLEA
GKFEDELQSVNGARNFVLLVSPGALDRCMQDHDCKDWVHKEIVTALSCG
ENYVPIIDGF
EWFPEQVLPEDMQAVLTFNGIKWSHEVQEAETIEKIIKFLQGRSSKRDSSAG
SDTSLGAAAP
MGPT
```



- 参数选择：Hphob./Kyte& Doolittle。横坐标是氨基酸的排列，纵坐标正半轴是疏水性，负半轴是亲水性程度。
- SARM1蛋白总体表现为亲水性。



预测SARM1蛋白信号肽

利用signal4.1 server分析

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

SignalP-5.0 Server

Version history [Click here to learn what is new and access previous versions.](#)

History paper [Click here to read "A Brief History of Protein Sorting Prediction", The Protein Journal, 2019](#)

Predict [Instructions/Help](#) [Data](#) [Abstract/Cite](#) [FAQ](#) [Portable version](#)

Submit data

The SignalP 5.0 server predicts the presence of signal peptides and the location of their cleavage sites in proteins from Archaea, Gram-positive Bacteria, Gram-negative Bacteria and Eukarya. In Bacteria and Archaea, SignalP 5.0 can discriminate between three types of signal peptides:

- Sec/SPI: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (Lep)
- Sec/SPII: lipoprotein signal peptides transported by the Sec translocon and cleaved by Signal Peptidase II (Lsp)
- Tat/SPI: Tat signal peptides transported by the Tat translocon and cleaved by Signal Peptidase I (Lep)

SignalP 5.0 is based on a deep convolutional and recurrent neural network architecture including a conditional random field.

Protein sequences should be not less than 10 amino acids. The maximum number of proteins is 5000.

```
>sp|Q6SZW1|SARM1_HUMAN Sterile alpha and TIR motif-containing protein 1 OS=Homo sapiens OX=9606
GN=SARM1 PE=1 SV=1
MVLTLILLSAYKLCRFFAMSGPRPGAERLAVPGPDGGGGTGPWWAAGGRGPREVSPGAGTE
VODALERALPELQQALSALQAGGARAVGAGLAEVFQVVEAWLLPAVGREVAQGLCDAI
RLDGGDLLLRLLQAPELETRVQAARLLEQILVAENRDRVARIGLVILNLAKEREPEVEL
ARSAVGILLEHMFHSEETCORLVAAGGLDAVLYWCRRTDPALLRHGALALGNLALHGGOA
VQRRMVEKRAAEWLFPLAFSKEDELLRLHACLAVLATNKEVEREVERSGTLALVEPLV
ASLDGPRFARCLVDASDTSQGRGPDQLGRVPLLDNSNRLEAQCIGAFYLCAAAIKSLQG
KTKVFSDIGAIQSLKRLVSYSTNGTKSALAKRALRLGEEVPRPILFSPVSWKEAEVQTV
IIOIGFSYDGIQSLKRLVSYSTNGTKSALAKRALRLGEEVPRPILFSPVSWKEAEVQTV
IIRTFIFI OTDI GMKSGITRKRFRFIFI TFI KTFANYS
```

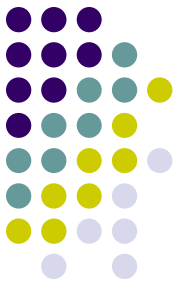
Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.09909	0.9091

Download: [PNG](#) / [EPS](#) / [Tabular](#)

SignalP-5.0 prediction (Eukarya): sp_Q6SZW1_SARM1_HUMAN

The graph displays three probability curves: SP (Sec/SPI) in orange, CS in green, and OTHER in blue. The SP curve starts at approximately 0.85 and remains high until the end of the protein. The CS and OTHER curves are very low, near zero, indicating a strong prediction of a signal peptide.

- 信号肽预测评分在0.09909，认为该蛋白不存在信号肽，是非分泌蛋白。



SARM1蛋白磷酸化位点预测

利用NetPhos 2.0分析

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

DTU Bioinformatics
Department of Bio and Health Informatics

Services are gradually being migrated to <https://services.healthtech.dtu.dk/>.
In the near future, cbs.dtu.dk will be retired. Please try out the new site.

[Home](#)

NetPhos 3.1 Server

The **NetPhos 3.1** server predicts serine, threonine or tyrosine phosphorylation sites in eukaryotic proteins using ensembles of neural networks. Both generic and kinase specific predictions are performed. The generic predictions are identical to the predictions performed by [NetPhos 2.0](#). The kinase specific predictions are identical to the predictions by [NetPhosK 1.0](#). Predictions are made for the following 17 kinases:

ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 and p38MAPK.

See the [version history](#) of this server.

NOTE: the online service at <http://www.cbs.dtu.dk/services/NetPhosK> is currently off-line; for the kinase specific predictions this service should be used instead.

Instructions	Output format	PhosphoBase
--------------	---------------	-------------

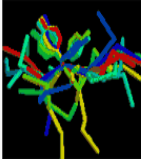
SUBMISSION

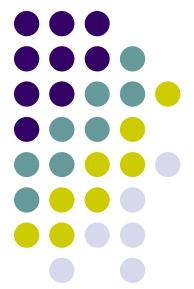
Paste a single sequence or several sequences in **FASTA** format into the field below:

```
GKPEDKLIQSYMGARNFVFLVSPGALDKCMQDHDCKDVVHKEIVTALSCGKNTVPIIDGF
EWPEPQVLPEDMQAVLTFMGIKWSHEYQEAETIEKIIIRFLQGKSSRDSSAGSDTSLBGAAP
MGPT
```

Submit a file in **FASTA** format directly from your local disk:

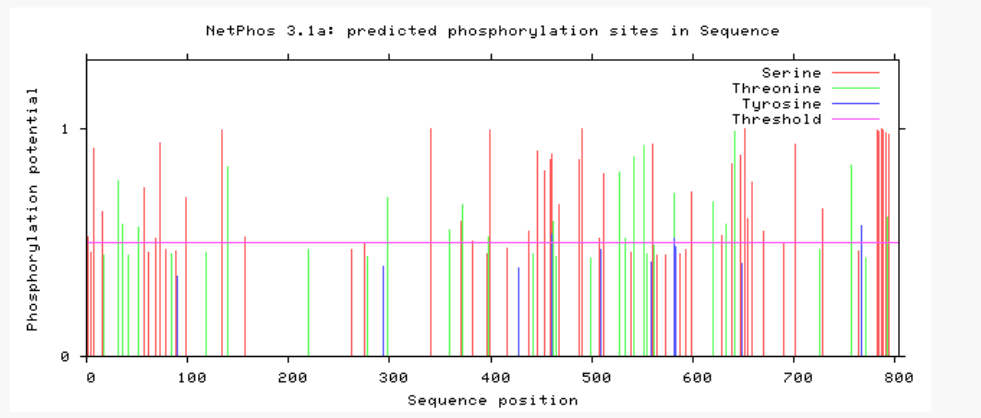
未选择任何文件



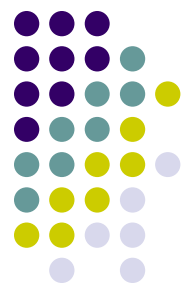


SARM1蛋白磷酸化位点预测

NSGSQLASLLKVLQLHGFVSVFIDVEKLEACKFEDKLIQSVMGARNFVLV	#	700
LSPGALDKCMQDHDCKDWVHKIEIVTALSCGKNIVPIIDGFEPQVLP	#	750
DMQAVLTFNGIKWSHEYQEATIEKIIIRFLQGRSSRDSSAGSDTSLEGAAP	#	800
MGPT	#	850
%1 S.....S.....S.....T...T.....	#	50
%1 .T...S.....S.....S.....	#	100
%1S.....T.....	#	150
%1S.....	#	200
%1S.....T..	#	250
%1S.....S.....	#	300
%1T.....S.T.....S.....TS.	#	350
%1S.....S.....S.....	#	400
%1 ..S...SYST...S.....S.S.....	#	450
%1S.....S.....T...T.....T.....	#	500
%1 T.....S.....YT.....S.....	#	550
%1T.....S...T...S..T...S...	#	600
%1 .S.S...S.....S.....	#	650
%1 .S.....S.....	#	700
%1T.....Y.....SS..SS..S.TS.....	#	750
%1	#	800



- 预测SARM1蛋白质中有39个Ser残基、19个Thr残基和3个Tyr残基可能成为磷酸化位点。该图为潜在的磷酸化位点分布。



分析SARM1蛋白糖基化位点

- NetOGlyc 4.0 Server:

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

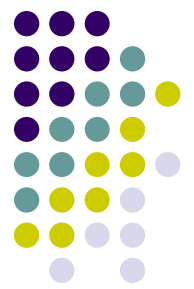
```

NetOGlyc 4.0 Server - prediction results
Technical University of Denmark

##ff-version 2
##source-version NetOGlyc 4.0.0.13
##date 19-12-20
##Type Protein
#seqname source feature start end score strand frame comment
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 4 4 0.0808596
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 8 8 0.0494479
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 19 19 0.17681
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 39 39 0.540034 #POSITIVE
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 54 54 0.907525 #POSITIVE
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 59 59 0.210959
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 77 77 0.331805
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 140 140 0.339362
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 183 183 0.245637
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 195 195 0.0605616
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 198 198 0.0819305
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 218 218 0.0375398
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 260 260 0.111639
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 279 279 0.0731779
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 290 290 0.0615579
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 292 292 0.11821
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 302 302 0.187126
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 316 316 0.267856
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 318 318 0.0893836
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 319 319 0.281675
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 336 336 0.132384
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 357 357 0.0377712
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 362 362 0.0604069
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 366 366 0.239937
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 373 373 0.197637
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 379 379 0.449627
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 381 381 0.642423 #POSITIVE
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 382 382 0.605978 #POSITIVE
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 385 385 0.594133 #POSITIVE
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 387 387 0.452786
SP_Q6SZW1_SARM1_HUMAN netOGlyc-4.0.0.13 CARBOHYD 408 408 0.482781
    
```

SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	408	408	0.482781	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	411	411	0.156403	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	419	419	0.126884	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	427	427	0.118643	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	432	432	0.0204182	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	447	447	0.10515	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	453	453	0.45658	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	459	459	0.166911	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	462	462	0.51709	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	471	471	0.239615	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	475	475	0.176525	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	480	480	0.129948	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	481	481	0.249317	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	485	485	0.226003	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	493	493	0.308134	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	502	502	0.273742	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	507	507	0.179625	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	513	513	0.0368179	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	519	519	0.467744	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	540	540	0.517058	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	548	548	0.509696	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	553	553	0.686127	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	558	558	0.738686	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	561	561	0.579692	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	567	567	0.627469	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	572	572	0.256896	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	574	574	0.23171	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	578	578	0.195613	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	590	590	0.123946	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	610	610	0.0321973	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	622	622	0.0420665	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	645	645	0.125417	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	648	648	0.0646333	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	677	677	0.117755	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	684	684	0.238444	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	691	691	0.063512	
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	703	703	0.601622	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	704	704	0.58496	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	707	707	0.82668	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	708	708	0.746079	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	711	711	0.112085	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	713	713	0.255478	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	714	714	0.798587	#POSITIVE
SP_Q6SZW1_SARM1_HUMAN	netOGlyc-4.0.0.13	CARBOHYD	724	724	0.106767	

- NetOGlyc 4.0 Server分析出SARM1蛋白有18个潜在的O-糖基化位点 (分值大于0.5)



分析SARM1蛋白糖基化位点

- NetNGlyc 1.0 Server:

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

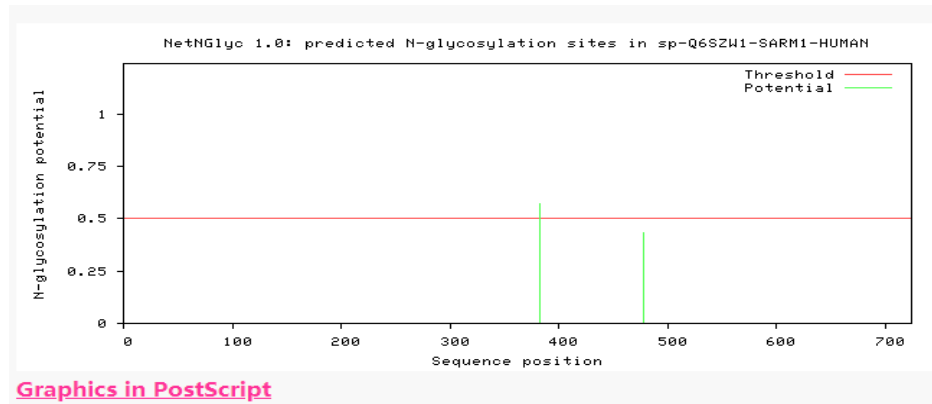
Asn-Xaa-Ser/Thr sequons in the sequence output below are highlighted in **blue**.
Asparagines predicted to be N-glycosylated are highlighted in **red**.

Output for 'sp_Q6SZW1_SARM1_HUMAN'

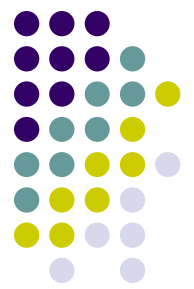
```
Name: sp_Q6SZW1_SARM1_HUMAN Length: 724
MVLILLLSAYKLCRFFAMSGPRPGAERLAVPGPDGGGIGPWWAAGGRPREVSPGAGTEVQDALERALPELQALSAK 80
QAGGARAVGAGLAEEVVFQVEEAWLLPAVGRVVAQGLCDARLDGGLDLLRLQLQAPELETRVQAARLLEQILVAENRDRV 160
ARIGLGVILHLAKEREPEVELARSVAGILEHMFKHSEETCQRLVAAGGLDAVLYWCRRTDPALLRHCALALGNALHGGQA 240
VQRRMVEKRAAEWLFPLAFSKEDELLRLHACLAVAVLATNKEVEREVERSGTLALVEPLVASLDPGRFARCLVDASDTSQ 320
GRGPDLDLQRLVPLLDNRLEAQCIGAFYLCAAEAIKSLQGGKTRVFDIGATQSLKRLVSYSTGTRKSAKAKRALRLGEE 400
VPRPILPVPVSWKAEAVGVLWQQIGFVKYCFSEFRQQVDGDLRLLEELQDPLGMRSGITRKRFRRELTELKTFANYS 480
ICDRSHLADWLGSLDPRFRQYTYGLVSCCLDRSLHRYSEQQLLEDCGTHLGVHRAKILIAAREMLHSPLPCTGGKPSGD 560
TPDVFISYRRRNSGSQLASLLKVLHLQLHGFVFDVVEKLEACKFEDKLIQSVMGARHFVLYVSPGALDKCMQDHDCKDWFH 640
KEIVTALSCGKNIVPILIDGFEWPEPQVLPEDMQAVLTFNGIKWSHEYQEATIEKIIRFLQGRSSRDSSAGSDTSLGGAAP 720
MGPI 800
```

(Threshold=0.5)

SeqName	Position	Potential	Jury agreement	N-Glyc result
sp_Q6SZW1_SARM1_HUMAN	383	NGTK	0.5694	(6/9) +
sp_Q6SZW1_SARM1_HUMAN	478	NYST	0.4308	(6/9) -



- NetNGlyc 1.0 Server分析出蛋白有1个潜在的N-糖基化位点 (分值大于0.5)。



构建SARM1蛋白系统发生树

- 选取不同物种的SARM1蛋白构建系统发生树
 - 首先将人的SARM1蛋白序列找出，并寻找几个代表性物种：黑猩猩、金丝猴和豚尾猴以及小鼠和大鼠的SARM1蛋白序列，一并下载

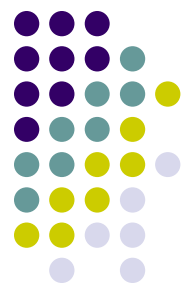
Filter by: BLAST Align Download Add to basket Columns > 1 to 25 of 1,257 Show

Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/> Q6PDS3	SARM1_MOUSE	NAD(+) hydrolase SARM1	Sarm1 KIAA0524	Mus musculus (Mouse)	724
<input type="checkbox"/> Q6SZW1	SARM1_HUMAN	NAD(+) hydrolase SARM1	SARM1 KIAA0524, SAMMD2, SARM	Homo sapiens (Human)	724
<input type="checkbox"/> I3L5V6	SARM1_PIG	NAD(+) hydrolase SARM1	SARM1	Sus scrofa (Pig)	725
<input type="checkbox"/> Q6IDD9	SARM1_DROME	NAD(+) hydrolase sarm1	Ect4 dsarm, CG43119	Drosophila melanogaster (Fruit fly)	1,360
<input type="checkbox"/> Q86DA5	SARM1_CAEEL	NAD(+) hydrolase tir-1	tir-1 nsy-2, F13B10.1	Caenorhabditis elegans	1,000
<input type="checkbox"/> F1QWA8	SARM1_DANRE	NAD(+) hydrolase SARM1	sarm1	Danio rerio (Zebrafish) (Brachydanio rerio)	713
<input type="checkbox"/> D3ZUM2	D3ZUM2_RAT	Sterile alpha and TIR motif contain...	Sarm1 Sarm1_predicted, rCG_33312	Rattus norvegicus (Rat)	724
<input type="checkbox"/> K7AVX3	K7AVX3_PANTR	SARM1 isoform 2	SARM1 CK820_G0028416	Pan troglodytes (Chimpanzee)	724
<input type="checkbox"/> H2NT32	H2NT32_PONAB	SARM1 isoform 2	SARM1 CR201_G0033492	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	724

Popular organisms: Human (19), Fruit fly (7), Mouse (7), C. elegans (6), Bovine (4)

Other organisms: Go

Search terms: Filter "sarm1" as: gene name (135)



构建SARM1蛋白系统发生树

- 利用phylip软件构建系统发生树：

The image shows two overlapping windows. The left window is ClustalX 2.1, displaying a multiple sequence alignment of SARM1 protein sequences from various species: 1_Rhinopithecus, SARM1_Macaca, SARM1_Human, ARM1_Chimpanzee, SARM1_Mouse, and SARM1_Rat. The alignment is color-coded by column. The right window is a terminal running 'seqboot.exe' from the phylib-3.695 package. The terminal output shows the following configuration and results:

```

C          Read categories of sites? No
S          Write out data sets or just weights? Data sets
I          Input sequences interleaved? Yes
0          Terminal type (IBM PC, ANSI, none)? IBM PC
1          Print out the data at start of run No
2          Print indications of progress of run Yes

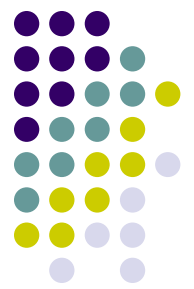
Y to accept these or type the letter for one to change
y

Random number seed (must be odd)?
5

completed replicate number 10
completed replicate number 20
completed replicate number 30
completed replicate number 40
completed replicate number 50
completed replicate number 60
completed replicate number 70
completed replicate number 80
completed replicate number 90
completed replicate number 100

Output written to file "outfile"
Done.
Press enter to quit.
  
```

首先用clustalX2.1进行多序列比对后保存为phy格式，然后打开phylib软件包，运行Seqboot、Prodist、Neighbor、Consense程序，得到邻接法产生的系统发生树。



构建SARM1蛋白系统发生树

```

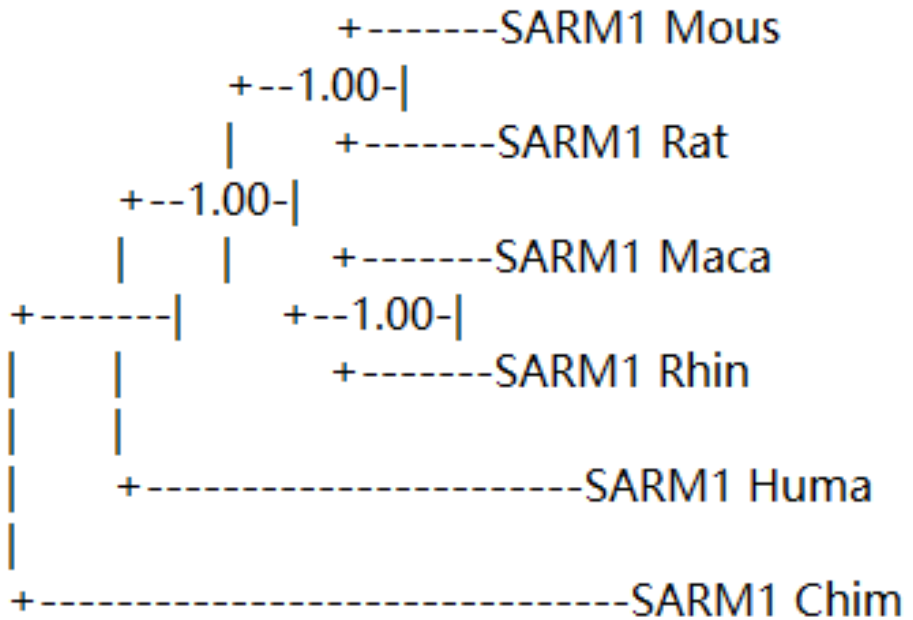
> phylip-3.695 > exe >
名称
drawtree.dll
DrawTree.jar
DrawTreeJava.exe
factor.exe
file1
file2
fitch.exe
font1
font2
font3
font4
font5
font6
gendist.exe
JavaPreview.ps
kitsch.exe
mix.exe
move.exe
neighbor.exe
outfile
outtree

C:\Users\16950\Desktop\phylip-3.695\exe\n
0      Outgroup
L      Lower-triangular data ma
R      Upper-triangular data ma
S      Subrepl
J      Randomize input order of spe
M      Analyze multiple data
0      Terminal type (IBM PC, ANSI, n
1      Print out the data at start o
2      Print indications of progress o
3      Print out
4      Write out trees onto tree

Y to accept these or type the lett
y

Cycle 3: species 5 ( 0.00841) jo
Cycle 2: species 1 ( 0.00001) jo
Cycle 1: node 1 ( 0.00286) joins
last cycle:
node 1 ( 0.00139) joins species

Output written on file "outfile"
Tree written on file "outtree"
Done.
Press enter to quit.
    
```



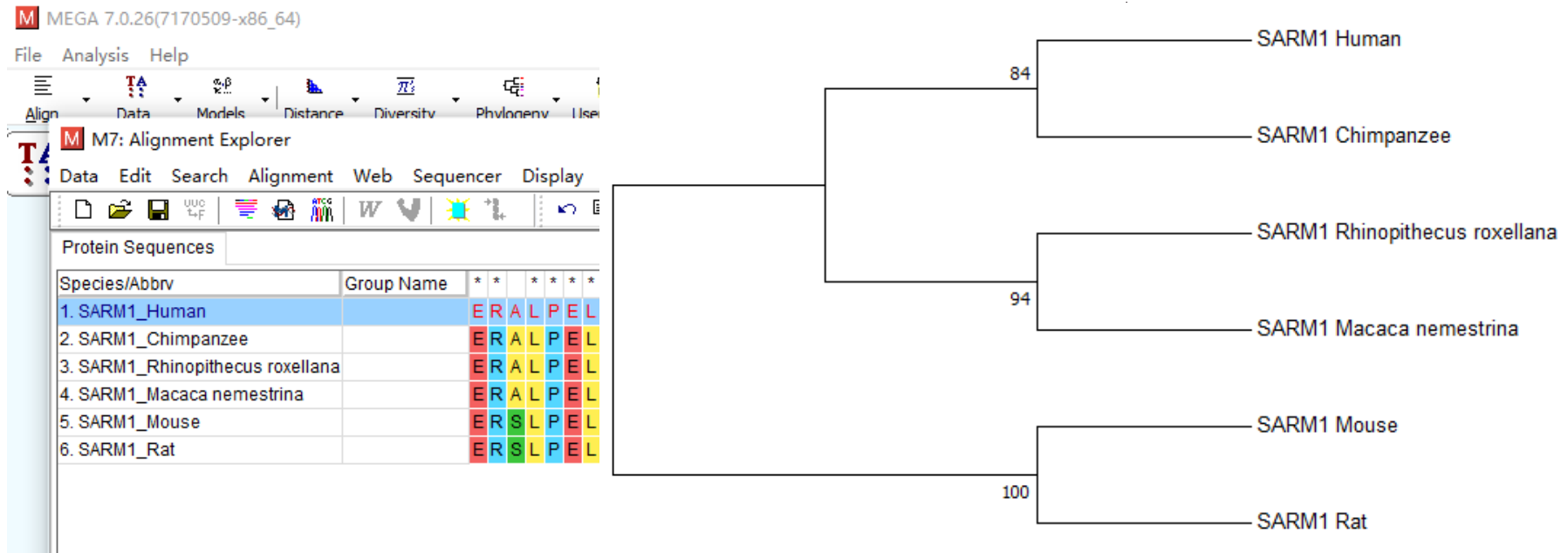
remember: this is an unrooted tree!

从phylip构建的系统发生树可以发现，小鼠和大鼠SARM1蛋白比较接近，金丝猴与豚尾猴比较接近，人和黑猩猩的SARM1蛋白比较接近。



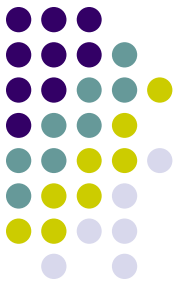
构建SARM1蛋白系统发生树

- 用MEGA7.0软件采用邻接法构建系统发生树：

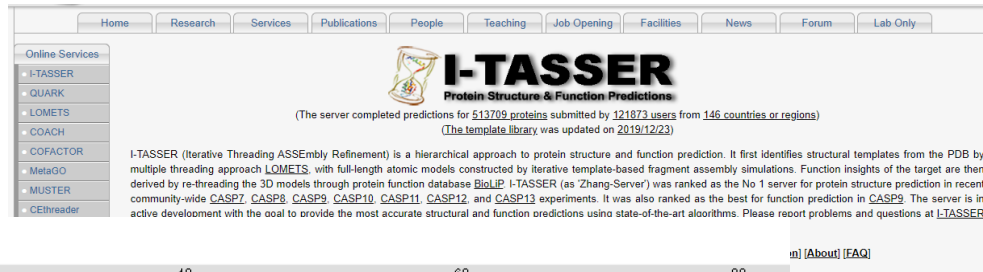


- 从SARM1蛋白系统发生树可以发现，人和黑猩猩的SARM1蛋白最为接近，其次是金丝猴与豚尾猴，更远一些的是小鼠和大鼠。

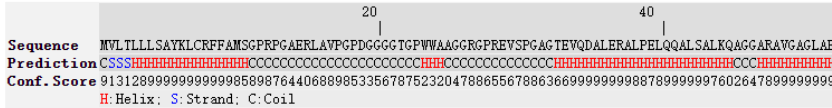
SARM1蛋白的三维结构预测



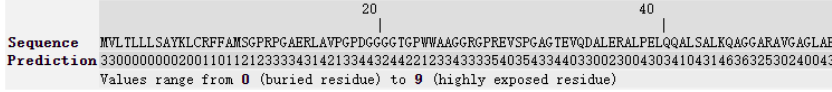
- I-TASSER: <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>



Predicted Secondary Structure

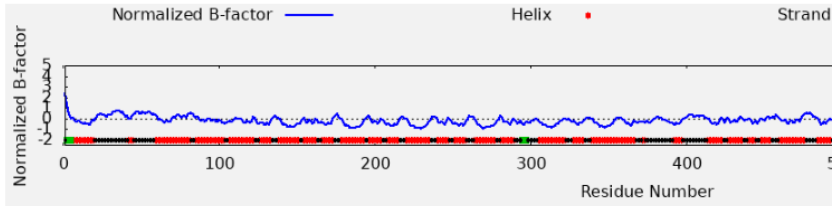


Predicted Solvent Accessibility



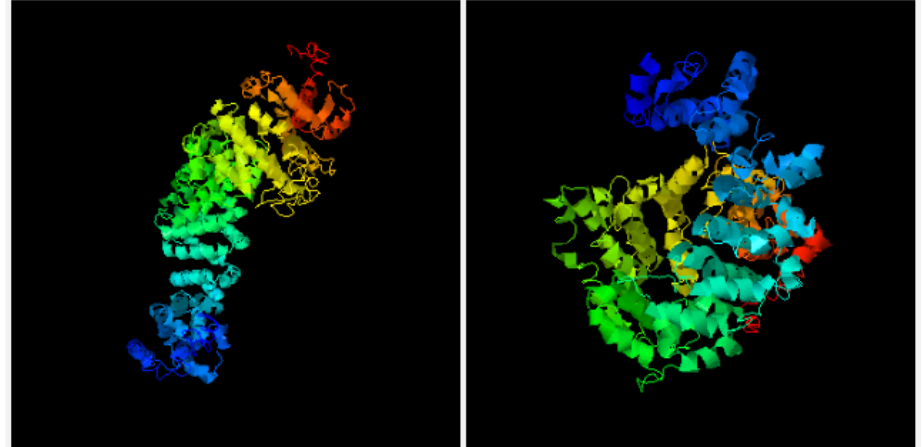
Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by $B = (B' - u)^2$, where B' is the raw B-factor value. [read more about predicted normalized B-factor](#))



- [More about C-score](#)
- [Local structure accuracy profile of the top five models](#)

(By right-click on the images, you can export image file or change the configurations, e.g. modifying the background color or stopping)

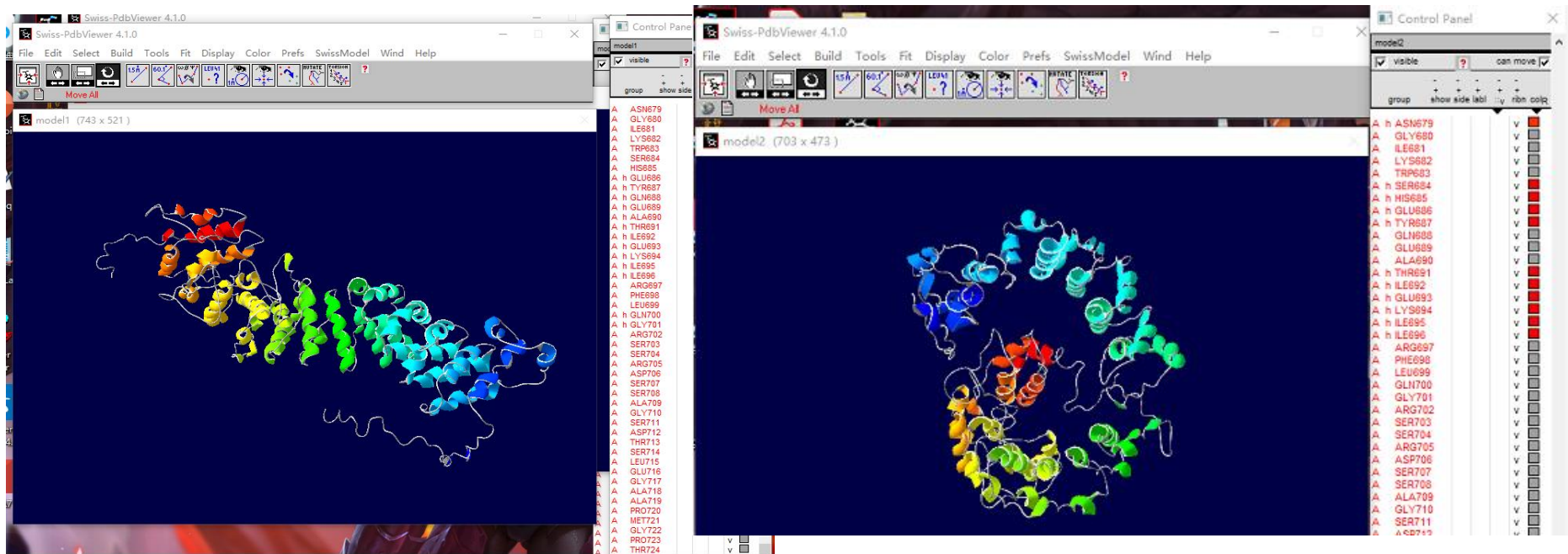


- Reset to initial orientation Spin On/Off
- [Download Model 1](#)
 - C-score = -2.36 ([Read more about C-score](#))
 - Estimated TM-score = 0.44 ± 0.14
 - Estimated RMSD = 14.0 ± 3.9 Å
 - [Download Model 2](#)
 - C-score = -1.91

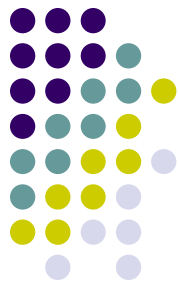


SARM1蛋白的三维结构预测

- 下载model1和model2, 用Swiss-PDB viewer打开



Model1显示SARM1的ARM、SAM以及TIR催化结构域是舒展的, 而model2是曲折的, SAM结构将TIR结构域包裹起来。说明model1的结果很大可能接近于SARM1活化后的结构, 而model2接近于SARM1自抑制的结构状态。



四、总结

- SARM1核酸分析
 - 通过ORF、PolyA位点、内含子/外显子剪切位点等分析，对SARM1基因结构更加清楚，有利于课题在基因水平对SARM1的表达和功能进行研究。
- SARM1蛋白分析
 - 通过分析了解SARM1蛋白的基本理化性质、二级结构、亲水性、亚细胞定位、磷酸化、糖基化位点、不同物种的亲缘性、可能的三维结构，对于课题研究SARM1蛋白潜在的功能和相关的信号通路有一定的参考意义。
- 课程总结
 - 提升了信息检索能力、开展蛋白研究的思维能力



谢谢大家!

附：小组分工情况

- 郭小春：核酸和蛋白结构分析
- 周梦婷：制作PPT、系统发生树构建
- 张钊祥：整理背景与总结
- 胡思梦：三维结构预测
- 江浩林、于雅馨、吴凡：协助分析、提出建议和修改PPT