

蛋白质性质和结构分析

日期: 2022-12-7

实验者: 生信 2001 张子栋

[MarkdownNotes/软件第8次作业.md at main · Bluuur/MarkdownNotes \(github.com\)](#)

[生物信息学原理/软件第8次作业.md · blur/MarkdownNotes - 码云 - 开源中国 \(gitee.com\)](#)

蛋白质性质和结构分析

实验目的

实验内容

实验步骤

Nanog

预测该基因编码产物的亚细胞定位。（至少使用两种预测方法，并比较不同方法的预测结果是否一致）

PSORT WoLF PSORT: Protein Subcellular Localization Prediction ([hgc.jp](#))

Yloc YLoc ([uni-tuebingen.de](#))

DeepLoc

人 Nanog 基因产物是否是糖蛋白？什么类型的糖蛋白？

GlycoEP Prediction of Glycosylation sites in Eukaryotic Proteins ([osdd.net](#))

NetOGlyc NetOGlyc - 4.0 - Services - DTU Health Tech

NetNGlyc Services- DTU Health Tech

分析人 Nanog 基因产物的亲水性和疏水性，列出最亲水和最疏水的位点。

分析甲虫基因 [AF422804](#)

等电点是多少？分子量是多少？是否含有 pfam 保守结构域？如有，列出登录号。

ExPASy - ProtParam tool

等电点

分子量

保守结构域

是否膜蛋白质？如果是膜蛋白质，请注明跨膜结构位点。是否具有 GPI 固定（anchor）的蛋白质？

Services- DTU Health Tech

预测该蛋白序列的二级结构，并简述预测结果（至少使用两种预测方法，并比较不同方法的预测结果是否一致）。

JPred JPred: A Protein Secondary Structure Prediction Server ([dundee.ac.uk](#))

PSIRPRED PSIPRED Workbench ([ucl.ac.uk](#))

PredictProtein Welcome to PredictProtein!

使用 Swiss-Model 预测该蛋白质序列的三级结构，并简述预测结果；该蛋白在 AlphaFoldDB 中是否有预测结果？如有，简述预测结果。

SWISS-MODEL Interactive Workspace ([expasy.org](#))

SWISS-MODEL 结果分析

Alpha Fold

Alpha Fold 结果分析

讨论

Nanog

分析甲虫基因 [AF422804](#)

实验目的

- 熟悉蛋白质序列和结构的主要分析内容。
- 在实践中逐步理解蛋白质序列和结构的主要分析算法的基本原理。

实验内容

- 综合使用多种在线工具，对蛋白质的一级、二级和三级结构进行分析和预测。
- 综合使用多种在线工具，对蛋白质的跨膜结构、翻译后修饰、亚细胞定位等进行分析 and 预测。

实验步骤

Nanog

1. Nanog 是 2003 年 5 月发现的一种转录因子，是一个有助于胚胎干细胞自我更新的关键因子，被认为在胚胎干细胞的全能性维持中起关键作用。人源 Nanog 基因，通常写成 Nanog1，位于 12 号染色体上。针对该基因（AY230262），请完成以下分析：

预测该基因编码产物的亚细胞定位。（至少使用两种预测方法，并比较不同方法的预测结果是否一致）

检索 AY230262 蛋白质序列 [NANOG [Homo sapiens](#)] - Protein - NCBI ([nih.gov](#))

```
1 >AAP49529.1 NANOG [Homo sapiens]
2 MSVDPACQSLPCFEASDCKESSMPVICGPEENYPQLQSSAEMPHTETVSPLPSSMDLLIQDSPDSST
3 SPKGGKQPTSAEKSAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQKYLQLQMQELSNILNLSYKQVK
4 TWFQNQRMKSKRWQKNNWPKNNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQT
5 QNIQSWNSHNSWNTQTWCTQSWNNQAWNPFYNGEESLQSCMQFQPNPASDLEAALEAAGEGLNVIQQT
6 TRYFSTPQTMDLFLNYSMMNPEDV
```

The screenshot displays the NCBI Protein database entry for NANOG [Homo sapiens]. The header shows the NIH logo and the text 'National Library of Medicine National Center for Biotechnology Information'. The protein name 'NANOG [Homo sapiens]' is prominently displayed. Below it, the GenBank accession number 'AAP49529.1' is shown. The protein sequence is presented in FASTA format, starting with '>AAP49529.1 NANOG [Homo sapiens]'. The sequence is: MSVDPACQSLPCFEASDCKESSMPVICGPEENYPQLQSSAEMPHTETVSPLPSSMDLLIQDSPDSST SPKGGKQPTSAEKSAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQKYLQLQMQELSNILNLSYKQVK TWFQNQRMKSKRWQKNNWPKNNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQT QNIQSWNSHNSWNTQTWCTQSWNNQAWNPFYNGEESLQSCMQFQPNPASDLEAALEAAGEGLNVIQQT TRYFSTPQTMDLFLNYSMMNPEDV. The page also features a 'Send to' button, a 'Change region shown' dropdown, and several analysis tools: 'Analyze this sequence' (Run BLAST), 'Identify Conserved Domains', and 'Highlight Sequence Features'. There are also links to 'Articles about the NANOG gene' and 'Reference sequence information'.

PSORT [WoLF PSORT: Protein Subcellular Localization Prediction \(hgcp.jp\)](#)

- 粘贴序列并提交

WoLF PSORT

Protein Subcellular Localization Prediction

[about WoLF PSORT](#) [WoLF PSORTについて](#) [links](#) [Example Output](#)

Please select an organism type:

☒ Animal

☐ Plant


☐ Fungi

Please select input method:

☒ From Text Area

☐ From File

Input Filename: 未选择文件



Text Area: Enter multifasta format protein sequence(s) here.

```
MSVDPACPQSLPCFEASDCKESSMPVIGPEENYPQLQMSSAE
MPHTETVSPLPSSMDLLIQDSPDSSTSPKQKQPTSAEKSVAKKEDKVPVKKQKTRTVF
SSTQLCVLNDRFQRQKYLSQLQMQLSNILNLSYKQVKTWFQNRMKSKRWQKNNWPK
NSNGVTQKASAPTYPSTLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTNISWSN
HSWNTQTWCTQSWNNQAWNPFYNCGEESLQSCMQFQPNSPASDLEAAEAAGEGLNV
IQQTRYFSTPQTMDLFLNYSMMNPEDV
```

(TSelect organism type to activate the submit button)


Contact:
[Functional Analysis in silico | NAKAI Lab](#)
8F General Research Bldg., 4-6-1 Shirokanedai Minato-ku Tokyo 108-8639, Japan
E-mail: **fais "AT" hgc.jp** (relace "AT" to @)

- 预测结果为 `nuc1`，对照 WoLF PSORT 细胞定位缩写表，即亚细胞定位为细胞核。

queryProtein WoLFPSORT prediction nucl: 31.5, cyto_nucl: 16.5					
PSORT features and traditional PSORTII prediction					
32 Nearest Neighbors					
id	site	distance	identity	comments	
MYF6_CHICK	nucl	832.326	12.7869%	[Uniprot]	SWISS-PROT45:Nuclear.
HXB4_CHICK	nucl	956.031	12.7451%	[Uniprot]	SWISS-PROT45:Nuclear.
KLF6_HUMAN	nucl	1116.8	12.1311%	[Uniprot]	SWISS-PROT45:Nuclear. GO:0005653; C:perinuclear space; Evidence:TAS.
HXD4_CHICK	nucl	1139.3	10.8197%	[Uniprot]	SWISS-PROT45:Nuclear.
HXDB_CHICK	nucl	1203.24	11.0749%	[Uniprot]	SWISS-PROT45:Nuclear.
HXD9_MOUSE	nucl	1254.99	12.6844%	[Uniprot]	SWISS-PROT45:Nuclear.
MYF6_HUMAN	nucl	1368.79	10.8197%	[Uniprot]	SWISS-PROT45:Nuclear. GO:0005634; C:nucleus; Evidence:TAS.
MYF6_MOUSE	nucl	1424.4	11.8033%	[Uniprot]	SWISS-PROT45:Nuclear.
MYF6_RAT	nucl	1424.82	11.8033%	[Uniprot]	SWISS-PROT45:Nuclear.
HXB4_FUGRU	nucl	1493.11	12.7869%	[Uniprot]	SWISS-PROT45:Nuclear.
HXBD_MOUSE	nucl	1500.92	12.1311%	[Uniprot]	SWISS-PROT45:Nuclear.
HXD9_HUMAN	nucl	1728.39	10.2339%	[Uniprot]	SWISS-PROT45:Nuclear.
HXDB_NOTVI	nucl	1731.37	10.0977%	[Uniprot]	SWISS-PROT45:Nuclear.
HXA4_MORSA	nucl	1974.43	11.8033%	[Uniprot]	SWISS-PROT45:Nuclear.
GA1B_XENLA	nucl	1986.65	15.6593%	[Uniprot]	SWISS-PROT45:Nuclear.
HXBD_HUMAN	nucl	2045.47	12.7869%	[Uniprot]	SWISS-PROT45:Nuclear.
TWST_XENLA	nucl	2076.19	14.7541%	[Uniprot]	SWISS-PROT45:Nuclear.
P53_HUMAN	cyto_nucl	2078.39	13.7405%	[Uniprot]	SWISS-PROT45:Cytoplasmic and nuclear. GO:0005634; C:nucleus; Evidence:TAS.
HXB4_HUMAN	nucl	2103.22	11.1475%	[Uniprot]	SWISS-PROT45:Nuclear. GO:0005634; C:nucleus; Evidence:NAS.
HXDD_BRARE	nucl	2113.43	12.459%	[Uniprot]	SWISS-PROT45:Nuclear.
MUSC_MOUSE	nucl	2128.3	13.1148%	[Uniprot]	SWISS-PROT45:Nuclear.
HXA2_CHICK	nucl	2137.84	19.7333%	[Uniprot]	SWISS-PROT45:Nuclear.
HXAD_CHICK	nucl	2146.48	10.7492%	[Uniprot]	SWISS-PROT45:Nuclear.
GAT2_HUMAN	nucl	2197.46	15%	[Uniprot]	SWISS-PROT45:Nuclear. GO:0005634; C:nucleus; Evidence:TAS.
MYOD_CHICK	nucl	2205.62	11.8971%	[Uniprot]	SWISS-PROT45:Nuclear.
HXD4_MOUSE	nucl	2206.84	12.459%	[Uniprot]	SWISS-PROT45:Nuclear.

Yloc [Yloc \(uni-tuebingen.de\)](http://uni-tuebingen.de)

- 粘贴序列，选择预测模型、物种后进行预测。



Interpretable Subcellular Localization Prediction

Predict with YLoc

Tutorial

Information

YLoc is an interpretable prediction system for protein subcellular localization prediction. In addition to the predicted location, YLoc gives a reasoning why this prediction was made and which biological properties of the protein sequence lead to this prediction. Moreover, a confidence estimate helps users to rate predictions as trustworthy. YLoc+ is able to predict the location of multiple-targeted proteins with high accuracy.

Please copy your protein sequence(s) in one letter code in the box below. A single sequence can be either raw or in FASTA format, whereas multiple sequences need to be in FASTA format. As an alternative, you can upload a FASTA file.

Click here for example sequences: U13-HTXT(Q75W(GZ)) SUC2(P00724) AGT1(P21549) GLR1(P41921) FH(P07954)

MSVDPACQSLPCFEASDCKESSPMIVICGREENYPSLQMSAE
MPHTEYSPFSSMDLIIDQSPDSSTPAKGQTSAEKSVAKKEDKVPVKKQKTRTVF
SSTQLCVLNDRFQKQYLSLQMQELSNILNLSYKQVKTWTFQNRQMSKRWQKNWPK
NSNGVTQKASAPTPSYSLYHGGCLVNPTGNLPMWSNQTNWNSWNSQTNQIQSWSN
HSWNTQTWTCQSWNKQWNSPFINCEESLQSCMQFQPNVSPASDLEAALEAAGEGLNV
IQQTRYFSTPQIMDLFLNYSNMMPQEDV

OR upload a FASTA file

选择文件

未选择文件

Select prediction model: YLoc+

Version: Animals

Use GO term features: Yes

Predict


To view the results of a previous prediction enter the query ID.

Query ID:

View

- 亚细胞定位为细胞核。

QueryID: `ec8ab61daa7d0a3892ba1830b4268aa2`



Interpretable Subcellular Localization Prediction

Predict with YLoc

Tutorial

Information

Prediction Summary

Query ID: ec8ab61daa7d0a3892ba1830b4268aa2 Query Date: 25/12/2022 15:45:39
Prediction based on model YLoc-LowRes Animals.

The prediction of YLoc is displayed in the table below. Place the mouse cursor over the table to see more details. Click on the *Elucidate* button to obtain an explanation why this prediction has been made. [More Help](#)

Query Sequence	Predicted Location	Probability	Confidence	Detailed Info
AAP49529.1NANOG[HOMOSAPIENS]	Nucleus	98.78 %	strong (0.86)	Elucidate...

Why?

YLoc-LowRes Animals predicted that protein sequence AAP49529.1NANOG[HOMOSAPIENS] is located in the nucleus with a probability of 98.8%. YLoc has a strong confidence (0.86) that this prediction is reliable.

The most important reason for making this prediction is the nuclear typical GO term. 51% of the proteins from the nucleus have a similar attribute, whereas only about 2% of the proteins form the cytoplasm, mitochondrion, and secreted pathway show this property. Moreover, the protein has a present GO term GO:0005634 (nucleus). 76% of the proteins from the nucleus have a similar attribute, whereas only about 4% of the proteins form the mitochondrion and secreted pathway show this property. There are more properties that support the predicted location.

For more details click the *Elucidate* button.

DeepLoc

- 输入序列后提交。 [DeepLoc - 1.0 - Services - DTU Health Tech](#)

DeepLoc - 1.0

Prediction of eukaryotic protein subcellular localization using deep learning

DeepLoc-1.0 predicts the subcellular localization of eukaryotic proteins. It can differentiate between 10 different localizations: Nucleus, Cytoplasm, Extracellular, Mitochondrion, Cell membrane, Endoplasmic reticulum, Chloroplast, Golgi apparatus, Lysosome/Vacuole and Peroxisome.

NOTE: This is not the newest version of DeepLoc. To use the current version, please go to [DeepLoc 2.0](#).

Submission

Instructions

Data

Abstract

Downloads

Submission

Submit data

Paste or upload protein sequence(s) as fasta format to predict the subcellular localization. The prediction can take a few minutes per sequence.

Protein sequences should be not less than 10 and not more than 6000 amino acids.

```
>AAP49529.1 NANOG [Homo sapiens]
MTNPALQKQLPFASSKSSSPWPLCPPEENTPSLQMSABPHHTETVSPFPSMELLIQDPFREST
SPKGAQPTSAEVSVAKEEYVPVMAKTKTVFSSTQLVYLNDRFQQRKYLSLQMQGLSTLLNLSYHVK
TFEQMQKSKSQRQKNNPENSNGYTRQASAPTFPSLYSYHAGCLVNPTGNLPIHNSQTPWNSTNSQT
KVLKSSNSDHWYVETETGRNNAQNSPTVMGEESLQCMQKQSPASLEALLAAGEGLNLAQT
TRTFSTPTMQLFTYSINMKPEDM
```

For example proteins [Click here](#)

Format directly from your local disk: [选择文件](#) 未选择文件

Protein encoding:

- Profiles (accurate, 50 sequences maximum)
- BLOSUM62 (fast, 500 sequences maximum)

Submit

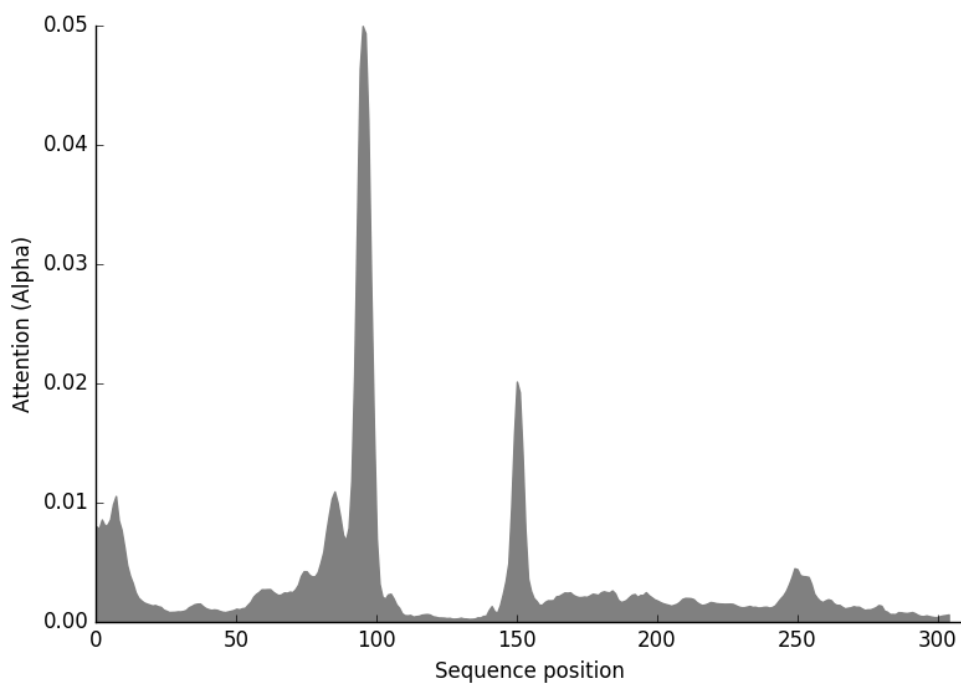
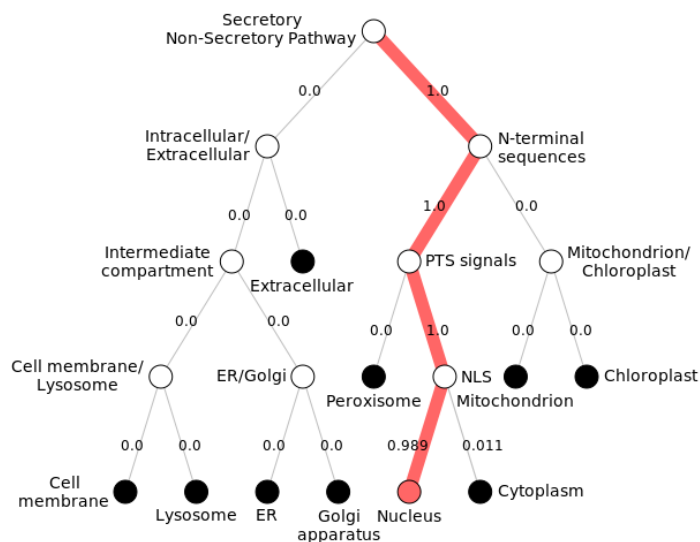
Clear fields

- 预测结果为细胞核。

Prediction: Nucleus, Soluble

Localization	Nucleus	Cytoplasm	Cell membrane	Endoplasmic reticulum	Golgi apparatus	Extracellular	Peroxisome	Mitochondrion	Lys
Likelihood	0.9924	0.0076	0	0	0	0	0	0	0

Type	Soluble	Membrane
Likelihood	0.877	0.123



[webface] Job finished



noreply@dtu.dk <noreply@dtu.dk>

23:50

收件人: zidongzh@outlook.com

Your job 63A870A8000051A120580292 has finished.

Details should be available at

<https://services.healthtech.dtu.dk/cgi-bin/webface2.cgi?jobid=63A870A8000051A120580292>

--DTU Health tech server queue

```

1      ``GFF
2      ##gff-version 2
3      ##source-version NetOGlyc 4.0.0.13
4      ##date 22-12-25
5      ##Type Protein
6      #seqname    source    feature start    end score    strand    frame    comment
7      AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    2      2      0.593531    .      .
      #POSITIVE
8      AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    10     10     0.35134     .      .
9      AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    17     17     0.124292    .      .
10     AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    22     22     0.397592    .      .
11     AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    23     23     0.601683    .      .
      #POSITIVE
12     AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    37     37     0.608048    .      .
      #POSITIVE
13     AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    41     41     0.823332    .      .
      #POSITIVE
14     AAP49529_1 netOGlyc-4.0.0.13    CARBOHYD    42     42     0.657648    .      .
      #POSITIVE

```

15	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	48	48	0.816338	.	.
	#POSITIVE						
16	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	50	50	0.787776	.	.
	#POSITIVE						
17	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	52	52	0.922385	.	.
	#POSITIVE						
18	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	56	56	0.860595	.	.
	#POSITIVE						
19	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	57	57	0.885228	.	.
	#POSITIVE						
20	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	65	65	0.899472	.	.
	#POSITIVE						
21	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	68	68	0.892244	.	.
	#POSITIVE						
22	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	69	69	0.991003	.	.
	#POSITIVE						
23	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	70	70	0.868445	.	.
	#POSITIVE						
24	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	71	71	0.982961	.	.
	#POSITIVE						
25	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	78	78	0.974264	.	.
	#POSITIVE						
26	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	79	79	0.990992	.	.
	#POSITIVE						
27	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	83	83	0.936976	.	.
	#POSITIVE						
28	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	98	98	0.311117	.	.
29	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	100	100	0.546425	.	.
	#POSITIVE						
30	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	103	103	0.138478	.	.
31	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	104	104	0.717357	.	.
	#POSITIVE						
32	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	105	105	0.126528	.	.
33	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	121	121	0.0999028	.	.
34	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	129	129	0.0246992	.	.
35	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	135	135	0.0108395	.	.
36	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	141	141	0.0336472	.	.
37	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	150	150	0.365898	.	.
38	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	162	162	0.724423	.	.
	#POSITIVE						
39	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	166	166	0.862944	.	.
	#POSITIVE						
40	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	170	170	0.797324	.	.
	#POSITIVE						
41	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	173	173	0.74059	.	.
	#POSITIVE						
42	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	176	176	0.768598	.	.
	#POSITIVE						
43	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	179	179	0.160221	.	.
44	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	180	180	0.482741	.	.
45	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	190	190	0.136736	.	.
46	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	197	197	0.169051	.	.
47	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	200	200	0.087622	.	.
48	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	204	204	0.143784	.	.
49	AAP49529_1 netOGlyc-4.0.0.13	CARBOHYD	205	205	0.0481689	.	.

50	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	207	207	0.113265	.	.
51	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	210	210	0.0859972	.	.
52	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	215	215	0.158727	.	.
53	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	217	217	0.189561	.	.
54	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	220	220	0.139987	.	.
55	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	223	223	0.0932478	.	.
56	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	225	225	0.0317189	.	.
57	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	228	228	0.117951	.	.
58	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	230	230	0.101636	.	.
59	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	238	238	0.0321494	.	.
60	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	247	247	0.0359253	.	.
61	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	250	250	0.0473348	.	.
62	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	258	258	0.224786	.	.
63	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	261	261	0.0486538	.	.
64	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	280	280	0.595211	.	.
	#POSITIVE							
65	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	281	281	0.660622	.	.
	#POSITIVE							
66	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	285	285	0.753219	.	.
	#POSITIVE							
67	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	286	286	0.733299	.	.
	#POSITIVE							
68	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	289	289	0.294352	.	.
69	AAP49529_1	netOGlyc-4.0.0.13	CARBOHYD	297	297	0.170629	.	.
70	...							

NetOGlyc-4.0 Server Output - DTU Health Tech

```
##gff-version 2
##source-version NetOGlyc 4.0.0.13
##date 22-12-25
##Type Protein
#seqname      source  feature start  end    score  strand  frame  comment
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 2      2      0.593531 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 10     10     0.35134 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 17     17     0.124292 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 22     22     0.397592 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 23     23     0.601683 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 37     37     0.608048 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 41     41     0.823332 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 42     42     0.657648 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 48     48     0.816338 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 50     50     0.787776 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 52     52     0.922385 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 56     56     0.860595 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 57     57     0.885228 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 65     65     0.899472 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 68     68     0.892244 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 69     69     0.991003 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 70     70     0.868445 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 71     71     0.982961 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 78     78     0.974264 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 79     79     0.990992 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 83     83     0.936976 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 98     98     0.311117 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 100    100    0.546425 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 103    103    0.138478 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 104    104    0.717357 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 105    105    0.126528 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 121    121    0.0999028 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 129    129    0.0246992 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 135    135    0.0108395 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 141    141    0.0336472 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 150    150    0.365898 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 162    162    0.724423 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 166    166    0.862944 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 170    170    0.797324 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 173    173    0.74059 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 176    176    0.768598 .      .      #POSITIVE
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 179    179    0.160221 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 180    180    0.482741 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 190    190    0.136736 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 197    197    0.169051 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 200    200    0.087622 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 204    204    0.143784 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 205    205    0.0481689 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 207    207    0.113265 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 210    210    0.0859972 .      .      .
AAP49529_1    netOGlyc-4.0.0.13 CARBOHYD 215    215    0.155737 .      .      .
```

[webface] Job finished



noreply@dtu.dk <noreply@dtu.dk>

0:07

收件人: zidongzh@outlook.com

Your job 63A8751500003926CFFCF00B has finished.

Details should be available at

<https://services.healthtech.dtu.dk/cgi-bin/webface2.cgi?jobid=63A8751500003926CFFCF00B>

--DTU Health tech server queue

NetNGlyc [Services- DTU Health Tech](#)

粘贴序列后提交。

```
#####
Warning: This sequence may not contain a signal peptide!!

Proteins without signal peptides are unlikely to be exposed to
the N-glycosylation machinery and thus may not be glycosylated
(in vivo) even though they contain potential motifs.

SignalP-NN euk predictions are as follows:

# name          Cmax   pos ?   Ymax   pos ?   Smax   pos ?   Smean ?   D       ?
AAP49529.1      0.108  62    0.101  37     0.107  56     0.095    0.098  N    0.450          SignalP-noTM

SignalP output is explained at https://services.healthtech.dtu.dk/services/SignalP-4.1/output.php
#####

Name: AAP49529.1      Length: 305
MSVDPACPQSLPCFEASDCKESSMPVICGPEENYPQLQSSAEMPHETVSPPLSSMDLLIQSPDSSTSPKCKQPTSA      80
EKSVAKKEDKVPVKQKQKTRTVFSSTQLCVLNDRFQRQKYLSQLQMQELSNILNLSYKQVKTWFPQNRMKSKRWQKNNWPK      160
YSNGVTTQKASAPTYPSLYSSYHQGLVNPPTGNLPWMSNQTWNNSTWSNQTNQISWSNHSWNTQTWCTQSWNNQANWSPF      240
YNCGEESLQSCMQFPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMMNQPEDV
.....N.....      80
.....N.....      160
.....N.....N.....      240
.....N.....      320

(Threshold=0.5)
-----
SeqName      Position  Potential   Jury      N-Glyc
              agreement result
-----
AAP49529.1   133 NLSY    0.5512     (7/9)     +
AAP49529.1   188 NPTG    0.5528     (7/9)     +   WARNING: PRO-X1.
AAP49529.1   198 NQTW    0.5532     (5/9)     +
AAP49529.1   202 NNST    0.4189     (6/9)     -
AAP49529.1   203 NSTW    0.4077     (7/9)     -
AAP49529.1   208 NQTQ    0.4834     (4/9)     -
AAP49529.1   218 NHSW    0.3536     (9/9)     --
AAP49529.1   295 NYSM    0.5649     (7/9)     +
```

Figure 2 is a bar chart showing the N-glycosylation potential for the protein sequence of the 120 kDa subunit of the 20S proteasome. The x-axis represents the sequence position from 0 to 300. The y-axis represents the N-glycosylation potential from 0 to 1. A horizontal purple line at y=0.5 indicates the threshold. Green vertical bars represent the potential at specific sequence positions. The bars are located at positions approximately 130, 185, 195, 200, 205, 210, 215, and 295.

- 分析人 Nanog 基因产物的亲水性和疏水性，列出最亲水和最疏水的位点。**

- ProtScale 粘贴序列后提交。 [ExPASy - ProtScale](#)

ExPASy

ProtScale

[Home](#) | [Contact](#)

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 scales and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program 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:

MSVDPACPQSLPCFEASDCKESSMPVICGPEENYPQLQSSAEMPHTETVSPLPSSMDLLIQDSPDSST
SPKGGKQPTSAEKSVAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQKYLSQLQMQELSNILNLSTYKQVK
TWFQNRMRKSKRWQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQT
QNIQSWNSHWNSTQWTCTQSWNNQAWNSPFYNCGEESLQSCMQFQNPSPASDLEAALEAAGEGLNVIQQT
TRYFSTPQTMDLFLNYSMMNPEDV

Please choose an amino acid scale from the following list. To display information about a scale (author, reference, amino acid scale values) you can click on its name.

☐ Molecular weight

☐ Bulkiness

☐ Polarity / Grantham

☐ Recognition factors

☐ Hphob. OM / Sweet et al.

☒ Hydropath. / Kyte & Doolittle

☐ Hphob. / Abraham & Leo

☐ Hphob. / Bull & Breese

☐ Hphob. / Guy

☐ Hphob. / Miyazawa et al.

☐ Hphob. / Roseman

☐ Hphob. / Wolfenden et al.

☐ Hphob. HPLC / Wilson & al.

☐ Hphob. HPLC pH5.4 / Cowan

☐ Hphob. / H₂ mobility

☐ HPLC / TFA retention

☐ HPLC / retention pH 2.1

☐ % buried residues

☐ Hphob. / Chothia

☐ Ratio hetero end/side

☐ % accessibility

☐ Number of codon(s)

☐ Polarity / Zimmerman

☐ Refractivity

☐ Hphob. / Eisenberg et al.

☐ Hphob. / Bopp & Woods

☐ Hphob. / Manavalan et al.

☐ Hphob. / Black

☐ Hphob. / Fauchere et al.

☐ Hphob. / Janin

☐ Hphob. / Ras & Argos

☐ Hphob. / Tanford

☐ Hphob. / Welling & al.

☐ Hphob. HPLC / Parker & al.

☐ Hphob. HPLC pH7.5 / Cowan

☐ HPLC / HPLC retention

☐ Transmembrane tendency

☐ HPLC / retention pH 7.4

☐ % accessible residues

☐ Hphob. / Rose & al.

☐ Average area buried

☐ alpha-helicit / Chou & Fasman

☐ beta-sheets / Chou & Fasman

- 最亲水位点: Position: 105 Score: 1.300 (max)
- 最疏水位点: Position: 153 Score: -3.156 (min)

1	MIN: -3.156, MAX: 1.300
2	Sequence:
3	MSVDPACPQSLPCFEASDCKESSMPVICGPEENYPQLQSSAEMPHTETVSPLPSSMDLLIQDSPDSST SPKGGKQPTSAEKSVAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQKYLSQLQMQELSNILNLSTYKQVK TWFQNRMRKSKRWQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQT QNIQSWNSHWNSTQWTCTQSWNNQAWNSPFYNCGEESLQSCMQFQNPSPASDLEAALEAAGEGLNVIQQT TRYFSTPQTMDLFLNYSMMNPEDV
4	
5	scale: Hydropathicity.
6	
7	window size: 9
8	scale not normalized.
9	Relative weight for window edges: 100 %
10	weight variation model: linear
11	
12	
13	Position: 5 Score: -0.067
14	Position: 6 Score: -0.367
15	Position: 7 Score: 0.144
16	Position: 8 Score: -0.500
17	Position: 9 Score: 0.167
18	Position: 10 Score: 0.656
19	Position: 11 Score: 0.067
20	Position: 12 Score: -0.011
21	Position: 13 Score: 0.078
22	Position: 14 Score: 0.078
23	Position: 15 Score: 0.444
24	Position: 16 Score: -0.411
25	Position: 17 Score: -0.622
26	Position: 18 Score: -0.989
27	Position: 19 Score: -1.389
28	Position: 20 Score: -1.178
29	Position: 21 Score: -1.167
30	Position: 22 Score: -1.256
31	Position: 23 Score: -0.400

32	Position:	24	Score:	-0.178
33	Position:	25	Score:	0.533
34	Position:	26	Score:	0.878
35	Position:	27	Score:	0.789
36	Position:	28	Score:	0.489
37	Position:	29	Score:	0.278
38	Position:	30	Score:	-0.322
39	Position:	31	Score:	-0.289
40	Position:	32	Score:	-0.933
41	Position:	33	Score:	-1.522
42	Position:	34	Score:	-1.378
43	Position:	35	Score:	-1.722
44	Position:	36	Score:	-1.333
45	Position:	37	Score:	-1.033
46	Position:	38	Score:	-0.733
47	Position:	39	Score:	-0.144
48	Position:	40	Score:	-0.389
49	Position:	41	Score:	-0.000
50	Position:	42	Score:	-0.089
51	Position:	43	Score:	-0.867
52	Position:	44	Score:	-0.556
53	Position:	45	Score:	-1.156
54	Position:	46	Score:	-1.144
55	Position:	47	Score:	-0.589
56	Position:	48	Score:	-0.878
57	Position:	49	Score:	-0.667
58	Position:	50	Score:	-0.456
59	Position:	51	Score:	-0.456
60	Position:	52	Score:	-0.189
61	Position:	53	Score:	-0.200
62	Position:	54	Score:	0.400
63	Position:	55	Score:	0.089
64	Position:	56	Score:	0.044
65	Position:	57	Score:	0.556
66	Position:	58	Score:	1.233
67	Position:	59	Score:	0.422
68	Position:	60	Score:	0.211
69	Position:	61	Score:	0.211
70	Position:	62	Score:	0.122
71	Position:	63	Score:	-0.478
72	Position:	64	Score:	-0.178
73	Position:	65	Score:	-0.689
74	Position:	66	Score:	-1.189
75	Position:	67	Score:	-1.778
76	Position:	68	Score:	-1.567
77	Position:	69	Score:	-1.611
78	Position:	70	Score:	-1.567
79	Position:	71	Score:	-1.822
80	Position:	72	Score:	-1.822
81	Position:	73	Score:	-1.911
82	Position:	74	Score:	-1.900
83	Position:	75	Score:	-1.911
84	Position:	76	Score:	-1.622
85	Position:	77	Score:	-1.833
86	Position:	78	Score:	-1.833

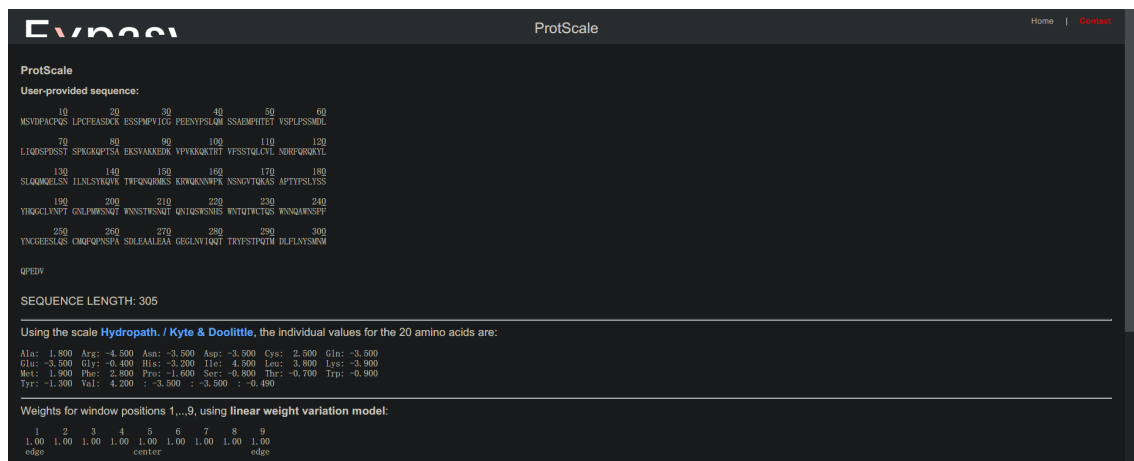
87	Position: 79	Score: -1.878
88	Position: 80	Score: -0.978
89	Position: 81	Score: -0.389
90	Position: 82	Score: -0.644
91	Position: 83	Score: -1.000
92	Position: 84	Score: -1.300
93	Position: 85	Score: -1.889
94	Position: 86	Score: -1.933
95	Position: 87	Score: -1.033
96	Position: 88	Score: -1.122
97	Position: 89	Score: -1.122
98	Position: 90	Score: -1.756
99	Position: 91	Score: -1.756
100	Position: 92	Score: -1.711
101	Position: 93	Score: -1.756
102	Position: 94	Score: -1.444
103	Position: 95	Score: -1.511
104	Position: 96	Score: -2.056
105	Position: 97	Score: -1.411
106	Position: 98	Score: -1.567
107	Position: 99	Score: -1.222
108	Position: 100	Score: -0.878
109	Position: 101	Score: -0.567
110	Position: 102	Score: -0.522
111	Position: 103	Score: -0.022
112	Position: 104	Score: 0.756
113	Position: 105	Score: 1.300 (max)
114	Position: 106	Score: 1.256
115	Position: 107	Score: 0.556
116	Position: 108	Score: 0.256
117	Position: 109	Score: -0.156
118	Position: 110	Score: 0.233
119	Position: 111	Score: 0.233
120	Position: 112	Score: -0.689
121	Position: 113	Score: -1.356
122	Position: 114	Score: -2.256
123	Position: 115	Score: -2.822
124	Position: 116	Score: -2.011
125	Position: 117	Score: -1.711
126	Position: 118	Score: -0.789
127	Position: 119	Score: -1.489
128	Position: 120	Score: -1.489
129	Position: 121	Score: -0.778
130	Position: 122	Score: -0.778
131	Position: 123	Score: -0.733
132	Position: 124	Score: -0.167
133	Position: 125	Score: -0.678
134	Position: 126	Score: -0.978
135	Position: 127	Score: -0.900
136	Position: 128	Score: -0.089
137	Position: 129	Score: -0.089
138	Position: 130	Score: 0.122
139	Position: 131	Score: 0.422
140	Position: 132	Score: 0.667
141	Position: 133	Score: -0.189

142	Position: 134	Score: -0.489
143	Position: 135	Score: 0.367
144	Position: 136	Score: -0.567
145	Position: 137	Score: -1.067
146	Position: 138	Score: -0.778
147	Position: 139	Score: -0.889
148	Position: 140	Score: -1.189
149	Position: 141	Score: -1.433
150	Position: 142	Score: -1.389
151	Position: 143	Score: -1.500
152	Position: 144	Score: -1.756
153	Position: 145	Score: -1.756
154	Position: 146	Score: -1.767
155	Position: 147	Score: -2.100
156	Position: 148	Score: -2.911
157	Position: 149	Score: -2.622
158	Position: 150	Score: -2.622
159	Position: 151	Score: -2.667
160	Position: 152	Score: -2.556
161	Position: 153	Score: -3.156 (min)
162	Position: 154	Score: -2.822
163	Position: 155	Score: -2.911
164	Position: 156	Score: -2.911
165	Position: 157	Score: -2.800
166	Position: 158	Score: -2.789
167	Position: 159	Score: -2.789
168	Position: 160	Score: -2.400
169	Position: 161	Score: -1.544
170	Position: 162	Score: -1.233
171	Position: 163	Score: -1.522
172	Position: 164	Score: -1.778
173	Position: 165	Score: -1.144
174	Position: 166	Score: -0.844
175	Position: 167	Score: -0.556
176	Position: 168	Score: -0.344
177	Position: 169	Score: -0.378
178	Position: 170	Score: -0.989
179	Position: 171	Score: -1.089
180	Position: 172	Score: -0.789
181	Position: 173	Score: 0.067
182	Position: 174	Score: -0.278
183	Position: 175	Score: -0.278
184	Position: 176	Score: -0.567
185	Position: 177	Score: -0.533
186	Position: 178	Score: -0.811
187	Position: 179	Score: -1.056
188	Position: 180	Score: -0.922
189	Position: 181	Score: -0.556
190	Position: 182	Score: -0.556
191	Position: 183	Score: 0.056
192	Position: 184	Score: -0.244
193	Position: 185	Score: -0.333
194	Position: 186	Score: -0.267
195	Position: 187	Score: 0.044
196	Position: 188	Score: 0.044

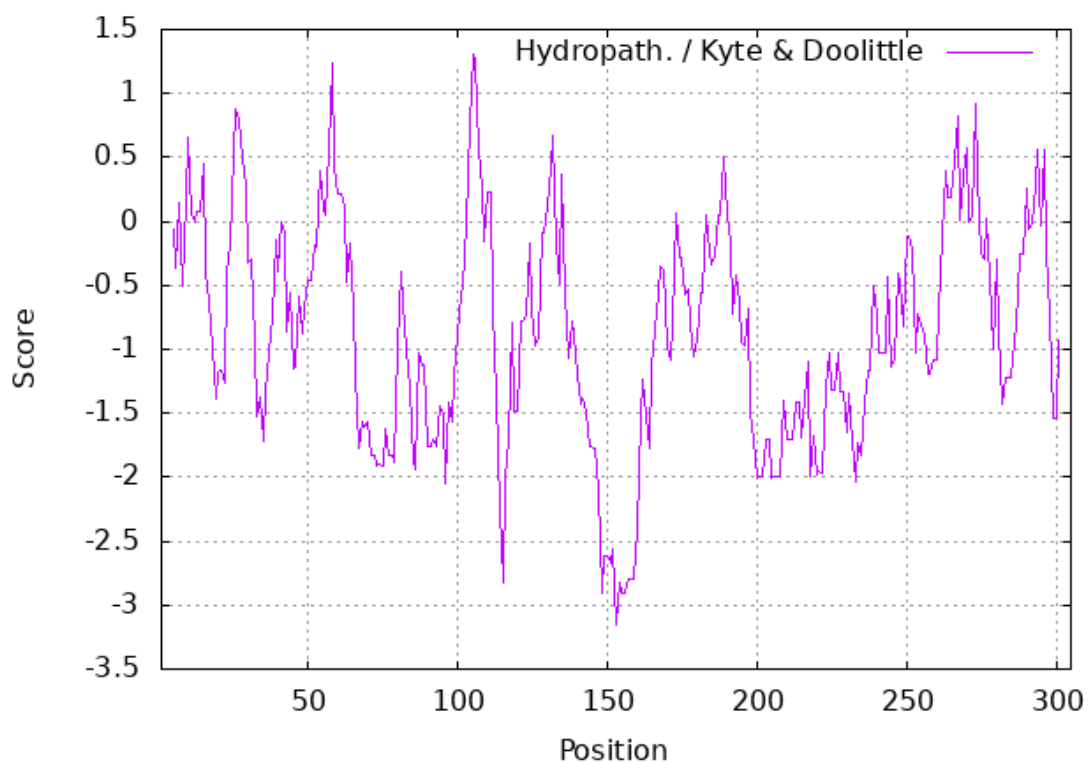
197	Position: 189	Score: 0.511
198	Position: 190	Score: 0.056
199	Position: 191	Score: -0.156
200	Position: 192	Score: -0.722
201	Position: 193	Score: -0.422
202	Position: 194	Score: -0.633
203	Position: 195	Score: -0.944
204	Position: 196	Score: -0.978
205	Position: 197	Score: -0.689
206	Position: 198	Score: -1.500
207	Position: 199	Score: -1.711
208	Position: 200	Score: -2.011
209	Position: 201	Score: -1.989
210	Position: 202	Score: -2.000
211	Position: 203	Score: -1.700
212	Position: 204	Score: -1.700
213	Position: 205	Score: -2.011
214	Position: 206	Score: -1.989
215	Position: 207	Score: -1.989
216	Position: 208	Score: -1.989
217	Position: 209	Score: -1.400
218	Position: 210	Score: -1.711
219	Position: 211	Score: -1.700
220	Position: 212	Score: -1.711
221	Position: 213	Score: -1.411
222	Position: 214	Score: -1.411
223	Position: 215	Score: -1.689
224	Position: 216	Score: -1.389
225	Position: 217	Score: -1.100
226	Position: 218	Score: -1.989
227	Position: 219	Score: -1.678
228	Position: 220	Score: -1.978
229	Position: 221	Score: -1.956
230	Position: 222	Score: -1.967
231	Position: 223	Score: -1.300
232	Position: 224	Score: -1.022
233	Position: 225	Score: -1.322
234	Position: 226	Score: -1.311
235	Position: 227	Score: -1.022
236	Position: 228	Score: -1.333
237	Position: 229	Score: -1.333
238	Position: 230	Score: -1.644
239	Position: 231	Score: -1.344
240	Position: 232	Score: -1.722
241	Position: 233	Score: -2.033
242	Position: 234	Score: -1.733
243	Position: 235	Score: -1.822
244	Position: 236	Score: -1.411
245	Position: 237	Score: -1.167
246	Position: 238	Score: -1.167
247	Position: 239	Score: -0.500
248	Position: 240	Score: -0.744
249	Position: 241	Score: -1.033
250	Position: 242	Score: -1.033
251	Position: 243	Score: -1.033

252	Position: 244	Score: -0.433
253	Position: 245	Score: -1.133
254	Position: 246	Score: -1.078
255	Position: 247	Score: -0.411
256	Position: 248	Score: -0.478
257	Position: 249	Score: -0.822
258	Position: 250	Score: -0.122
259	Position: 251	Score: -0.122
260	Position: 252	Score: -0.211
261	Position: 253	Score: -1.022
262	Position: 254	Score: -0.722
263	Position: 255	Score: -0.811
264	Position: 256	Score: -0.889
265	Position: 257	Score: -1.189
266	Position: 258	Score: -1.189
267	Position: 259	Score: -1.078
268	Position: 260	Score: -1.078
269	Position: 261	Score: -0.700
270	Position: 262	Score: -0.111
271	Position: 263	Score: 0.400
272	Position: 264	Score: 0.189
273	Position: 265	Score: 0.189
274	Position: 266	Score: 0.478
275	Position: 267	Score: 0.822
276	Position: 268	Score: 0.011
277	Position: 269	Score: 0.356
278	Position: 270	Score: 0.578
279	Position: 271	Score: -0.011
280	Position: 272	Score: 0.033
281	Position: 273	Score: 0.922
282	Position: 274	Score: 0.333
283	Position: 275	Score: -0.256
284	Position: 276	Score: -0.289
285	Position: 277	Score: 0.022
286	Position: 278	Score: -0.433
287	Position: 279	Score: -1.000
288	Position: 280	Score: -0.300
289	Position: 281	Score: -0.856
290	Position: 282	Score: -1.433
291	Position: 283	Score: -1.222
292	Position: 284	Score: -1.222
293	Position: 285	Score: -1.222
294	Position: 286	Score: -0.933
295	Position: 287	Score: -0.822
296	Position: 288	Score: -0.256
297	Position: 289	Score: -0.256
298	Position: 290	Score: 0.256
299	Position: 291	Score: -0.056
300	Position: 292	Score: -0.022
301	Position: 293	Score: 0.278
302	Position: 294	Score: 0.567
303	Position: 295	Score: -0.033
304	Position: 296	Score: 0.567
305	Position: 297	Score: -0.244
306	Position: 298	Score: -0.733

307	Position: 299	Score: -1.544
308	Position: 300	Score: -1.544
309	Position: 301	Score: -0.933



ProtScale output for user_sequence



分析甲虫基因 AF422804

分析一个甲虫基因（AF422804）编码的蛋白质的化学性质和结构特点（请注明分析方法名称）。

获取蛋白质序列 [ABC transmembrane transporter white [Tribolium castaneum](#)] - Protein - NCBI ([nih.gov](#))

```
1 >AAL56571.1 ABC transmembrane transporter white [Tribolium castaneum]
2 MENETEPLLSGVVSQINGNSGDSTSSATSIDLSTFRVPTYGTTSHPTSKLVPPDERITYSWTEINAFANV
3 SPPKTKFFNLIKRKDSPVQKKHILKNVFGVAYPGELLAILGSSGAGKTTLLNTLTFHTSSNLTVSGLRCV
4 NGIPVSSKTLASQSAYVQDDLFIGTLTVKEHLIFQALLRMDRDISYSQRMARVEEVISDLALSKCQNTP
5 IGILGRIKGISGGEKKRLSFAAEVLTNPKLMFCDEPTSGLDSFMALTVMQVLKEMAMTGKTVICTIHQPS
6 SEVYSMFDKLLLMSEGRATFLGSPEEAETFFRELEAPCPRNYNPADYFIQLLAIVPEKEESSRQAVNLIC
7 DKFERSNIGVKIALEAATTEREGGYHDIWMSGESFKSPYKASCWAQFKAVLWRSILAVFKEPLLIKVRLL
8 QTLIISLVIGAIYFGQDLNQDGMNINGVLFVFLTNMTFQNVFAVINVFSGELPVFLQEHNRNGMYRPSIY
9 FISKTLAESPIFIIIPVTLTSCYFMIGLNSHGFRFYIACGIMILVANVAISFGYLISCVSRVSMALSI
10 GPPLVIPFLLFGGFLLNVSSIPIYFKWLSFLSWFRYNGALMINQWENVNTNIQCPNADLPCPKDGHVILE
11 TFFHSEADFVMDVVMLAVLIVGFRFLVAFLALLVKTWRFK
```

The screenshot shows the NIH website interface for the protein entry 'ABC transmembrane transporter white [Tribolium castaneum]'. The main content area displays the protein sequence in FASTA format, along with GenBank accession number AAL56571.1 and links to GenPept, Identical Proteins, and Graphics. The right sidebar contains several sections: 'Analyze this sequence' with links to Run BLAST, Identify Conserved Domains, and Highlight Sequence Features; 'Articles about the W gene' with links to 'The ABCs of eye color in Tribolium castaneum: orthologs of the Drosophila white' and 'Genetic linkage maps of the red flour beetle, Tribolium castaneum, based on'; 'Reference sequence information' with a link to RefSeq protein; and 'More about the gene W' with a link to W gene. The bottom of the page features a footer with the NIH logo and contact information.

等电点是多少？分子量是多少？是否含有 pfam 保守结构域？如有，列出登录号。

Expsy - ProtParam tool

粘贴序列并提交。

The screenshot shows the Expsy ProtParam tool interface. The main content area contains a text input field for the protein sequence, a 'Compute parameters' button, and a 'RESET' button. The bottom of the page features a footer with the Expsy logo and contact information.

- 1 Number of amino acids: 669
- 2
- 3 Molecular weight: 74351.58

4

5 Theoretical pI: 7.52

6

7 Amino acid composition:

8	Ala (A)	41	6.1%
9	Arg (R)	25	3.7%
10	Asn (N)	31	4.6%
11	Asp (D)	21	3.1%
12	Cys (C)	12	1.8%
13	Gln (Q)	20	3.0%
14	Glu (E)	35	5.2%
15	Gly (G)	42	6.3%
16	His (H)	10	1.5%
17	Ile (I)	52	7.8%
18	Leu (L)	72	10.8%
19	Lys (K)	32	4.8%
20	Met (M)	20	3.0%
21	Phe (F)	46	6.9%
22	Pro (P)	32	4.8%
23	Ser (S)	61	9.1%
24	Thr (T)	40	6.0%
25	Trp (W)	8	1.2%
26	Tyr (Y)	18	2.7%
27	Val (V)	51	7.6%
28	Pro (O)	0	0.0%
29	Sec (U)	0	0.0%
30			
31	(B)	0	0.0%
32	(Z)	0	0.0%
33	(X)	0	0.0%
34			
35			
36	Total number of negatively charged residues (Asp + Glu): 56		
37	Total number of positively charged residues (Arg + Lys): 57		
38			
39	Atomic composition:		
40			
41	Carbon	C	3394
42	Hydrogen	H	5311
43	Nitrogen	N	855
44	Oxygen	O	952
45	Sulfur	S	32
46			
47	Formula: C3394H5311N855O952S32		
48	Total number of atoms: 10544		
49			
50	Extinction coefficients:		
51			
52	Extinction coefficients are in units of M ⁻¹ cm ⁻¹ , at 280 nm measured in water.		
53			
54	Ext. coefficient	71570	
55	Abs 0.1% (=1 g/l)	0.963, assuming all pairs of Cys residues form cystines	
56			
57			

58 Ext. coefficient 70820
59 Abs 0.1% (=1 g/l) 0.953, assuming all cys residues are reduced
60
61 Estimated half-life:
62
63 The N-terminal of the sequence considered is M (Met).
64
65 The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
66 >20 hours (yeast, in vivo).
67 >10 hours (Escherichia coli, in vivo).
68
69
70 Instability index:
71
72 The instability index (II) is computed to be 37.71
73 This classifies the protein as stable.
74
75
76
77 Aliphatic index: 100.52
78
79 Grand average of hydropathicity (GRAVY): 0.259

等电点

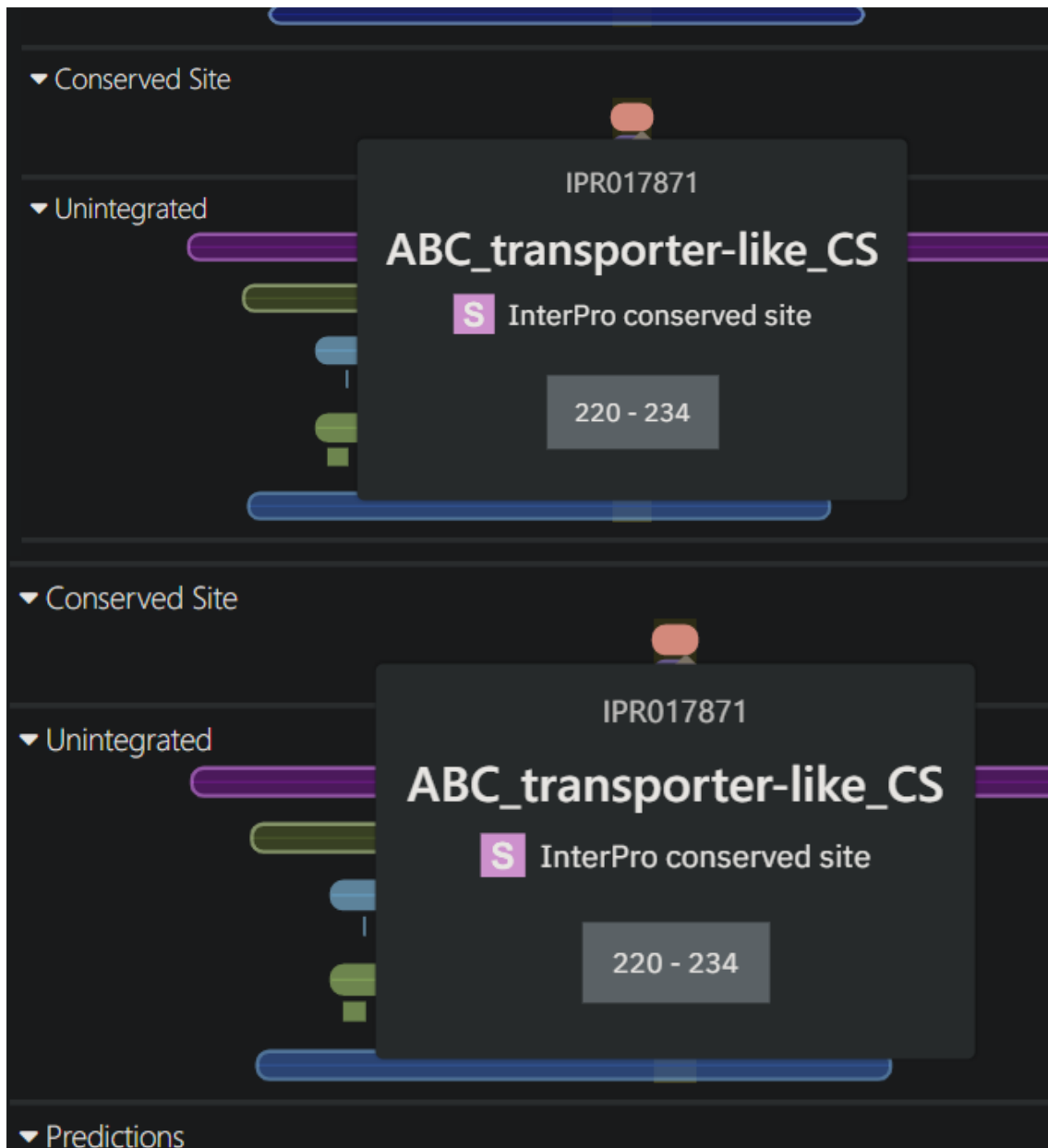
- Theoretical pI: 7.52

分子量

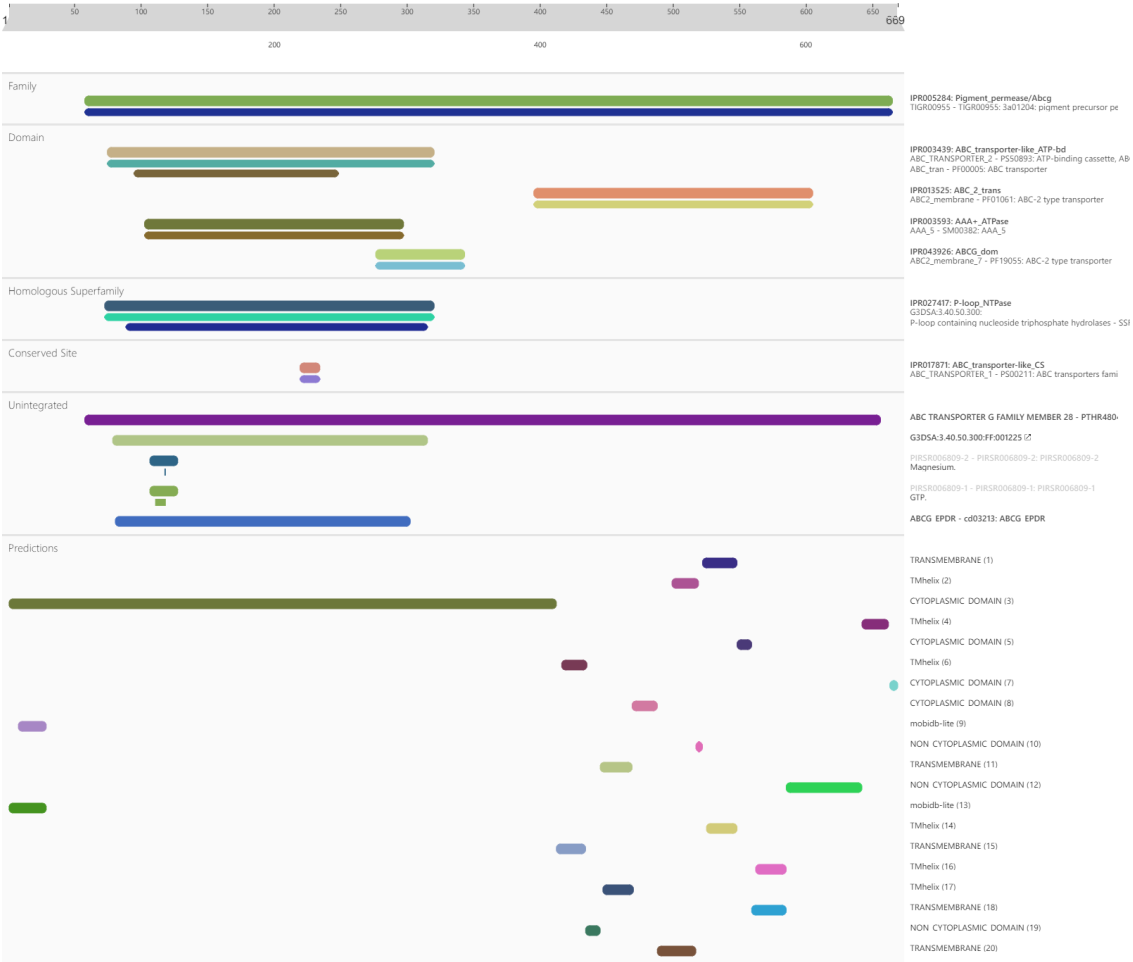
- Molecular weight: 74351.58

保守结构域

- The Pfam website will be decommissioned in January 2023. [Pfam is now hosted by InterPro \(xfam.org\)](#)
改用 InterProScan [InterPro \(ebi.ac.uk\)](#)



● 全部结果：



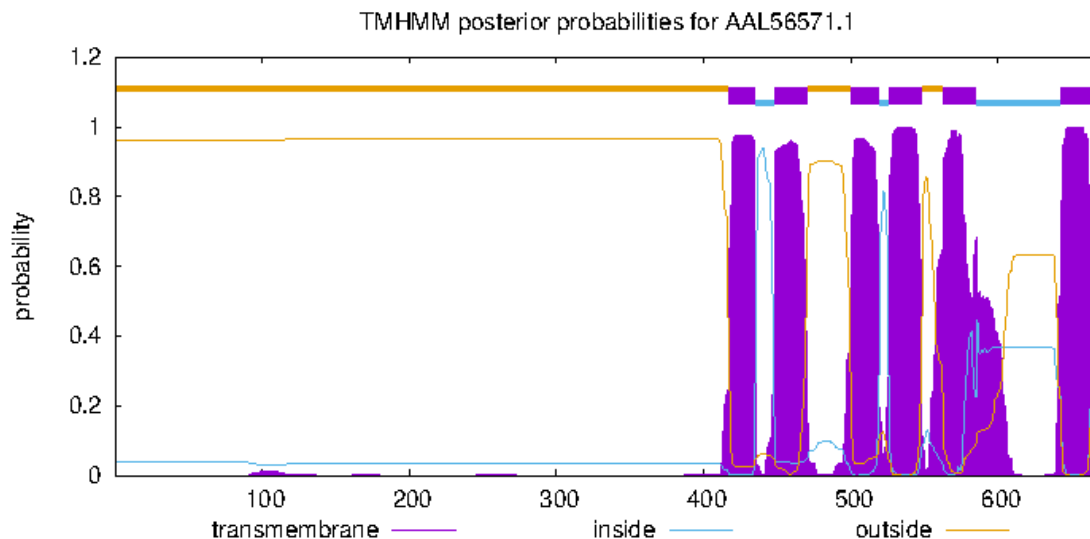
是否膜蛋白质？如果是膜蛋白质，请注明跨膜结构位点。是否具有 GPI 固定 (anchor) 的蛋白质？

[Services- DTU Health Tech](#)

粘贴序列并提交。

TMHMM result

```
# AAL56571.1 Length: 669
# AAL56571.1 Number of predicted TMHs: 6
# AAL56571.1 Exp number of AAs in TMHs: 138.42885
# AAL56571.1 Exp number, first 60 AAs: 8e-05
# AAL56571.1 Total prob of N-in: 0.03731
AAL56571.1 TMHMM2.0 outside 1 416
AAL56571.1 TMHMM2.0 TMhelix 417 435
AAL56571.1 TMHMM2.0 inside 436 447
AAL56571.1 TMHMM2.0 TMhelix 448 470
AAL56571.1 TMHMM2.0 outside 471 499
AAL56571.1 TMHMM2.0 TMhelix 500 519
AAL56571.1 TMHMM2.0 inside 520 525
AAL56571.1 TMHMM2.0 TMhelix 526 548
AAL56571.1 TMHMM2.0 outside 549 562
AAL56571.1 TMHMM2.0 TMhelix 563 585
AAL56571.1 TMHMM2.0 inside 586 642
AAL56571.1 TMHMM2.0 TMhelix 643 662
AAL56571.1 TMHMM2.0 outside 663 669
```

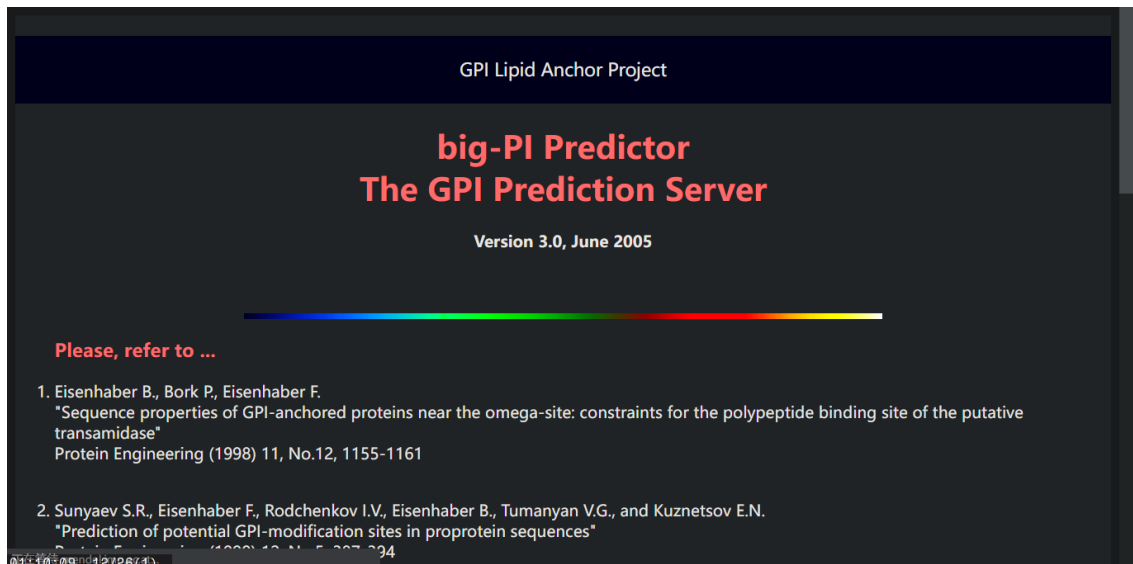


- 是膜蛋白，共有 6 个跨膜螺旋。
- 以下为跨膜位点

```
# AAL56571.1 Length: 669
# AAL56571.1 Number of predicted TMHs: 6
# AAL56571.1 Exp number of AAs in TMHs: 138.42885
# AAL56571.1 Exp number, first 60 AAs: 8e-05
# AAL56571.1 Total prob of N-in: 0.03731
AAL56571.1 TMHMM2.0 outside 1 416
AAL56571.1 TMHMM2.0 TMhelix 417 435
AAL56571.1 TMHMM2.0 inside 436 447
AAL56571.1 TMHMM2.0 TMhelix 448 470
AAL56571.1 TMHMM2.0 outside 471 499
AAL56571.1 TMHMM2.0 TMhelix 500 519
AAL56571.1 TMHMM2.0 inside 520 525
AAL56571.1 TMHMM2.0 TMhelix 526 548
AAL56571.1 TMHMM2.0 outside 549 562
AAL56571.1 TMHMM2.0 TMhelix 563 585
AAL56571.1 TMHMM2.0 inside 586 642
AAL56571.1 TMHMM2.0 TMhelix 643 662
AAL56571.1 TMHMM2.0 outside 663 669
```


- 分析 GPI 位点 [GPI Modification Site Prediction \(imp.ac.at\)](http://GPI.Modification.Site.Prediction.imp.ac.at)

粘贴序列并提交



- 是无 GPI 固定的蛋白质

Output of the prediction tool:

```

1  ~~~~~
2
3  Query sequence AAL56571.1 (length 669 amino acids):
4  MENETEPLLS GVVSQINGNS GDSTSSATSI DLSTFRVPTY GTTSHPTS KL VPPDERITYS
5  WTEINAFANV SPPKTKFFNL IKRKDSPVQK KHILKNVFGV AYPGELLAIL GSSGAGKTTL
6  LNTLTFTHTSS NLTVSGLRVC NGIPVSSKTL ASQSAYVQQD DLFIGTLTVK EHLIFQALLR
7  MDRDISYSQR MARVEEVISD LALSKCQNTF IGILGRIKGI SGGEKKRLSF AAEVLNPKL
8  MFCDEPTSLG DSFMALTMQ VLKEMAMTGK TVICTIHQPS SEVYSMFDKL LLMSEGRTAF
9  LGSPEEAETF FRELEAPCPR NYNPADYFIQ LLAIVPEKEE SSRQAVNLIC DKFERSNIGV
10 KIALEAATTE REGGYHDIWM SGESFKSPYK ASCWAQFKAV LWSILAVFK EPLLIKVRLL
11 QTLIISLVIG AIYFGQDLNQ DGVMNINGVL FVFLTNMTFQ NVFAVINVS GELPVFLQEH
12 RNGMYRPSIY FISKTLAESP IFIIIPVTLT SVCYFMIGLN SHGFRFYIAC GIMILVANVA
13 ISFGYLISCV SRSVSMALSI GPPLVIPFL FGGFFLNVS IPIYFKWLSF LSWFRYGNGA
14 LMINQWENVN NIQCPNADLP CPKDGHVILE TFHFSEADFV MDVVMLAVLI VGFRLLVAFLA
15 LLVKTWRFK
16
17 The site with the best score is shown in black.
18
19 ~~~~~
20 Prediction of potential C-terminal GPI-Modification Sites
21 ~~~~~
22
23 Use of the prediction function for METAZOA
24
25
26 None potential GPI-modification site was found.
27 Among all positions checked, sequence position 638 had the best score.
28
29
30 Total Score.....: -21.03
    (PValue = 2.419695e-02)
31 Components of the Score Function:
32 Profile Score.....: -9.72
33 Term 0 Contents and Windows of DE in Region [-11..1].....: 0.00

```

34	Term 1	Hydrophilicity of N-terminal Region [-11..1].....:	-0.00
35	Term 2	Penalty for low Profile Score in Region [0..2].....:	-4.00
36	Term 3	Volume Limitation [-1..+2].....:	-2.24
37	Term 4	Volume Compensation (-1, 1, 2).....:	-2.02
38	Term 5	Volume Compensation (-1, 2).....:	-0.02
39	Term 6	Backbone Flexibility [-1..2].....:	0.00
40	Term 7	Propeptide Length.....:	0.00
41	Term 8	Hydrophilicity of Spacer Region [3..8].....:	-0.48
42	Term 9	Volume Limitation [3..8].....:	-0.74
43	Term 10	Penalty for charged AAs in Spacer Region [3..10]....:	0.00
44	Term 11	Backbone Flexibility [3..8]....	0.00
45	Term 12	Penalty for low Profile Score in Region [10..end]..:	0.00
46	Term 13	Hydrophobicity of Tail [10..end].....:	-0.34
47	Term 14	Hydrophobicity of Tail [26..end].....:	-1.47
48	Term 15	Even Distribution of Hydrophobicity [9..end].....:	0.00
49	Term 16	Penalty for polar windows in Region [10..end].....:	0.00
50	Term 17	Penalty for SGC-windows in Region [10..end].....:	0.00
51	Term 18	LVI Contents [10..end].....:	0.00
52	Term 19	Penalty for FYHW - Sections in Region [10..end]....:	0.00
53	Term 20	Penalty for windows with small volume [10..end]....:	0.00
54		Profile independent Score.....:	-7.31

预测该蛋白序列的二级结构，并简述预测结果（至少使用两种预测方法，并比较不同方法的预测结果是否一致）。

JPred [JPred: A Protein Secondary Structure Prediction Server \(dundee.ac.uk\)](http://jpred.dundee.ac.uk)

1. 输入序列并提交。

Jpred 4
Incorporating Jnet

A Protein Secondary Structure Prediction Server

Home REST API About F.A.Q. Help & Tutorials Monitoring Contact Publications

Input sequence(?)


GPPLVIPFLFGGFFLVSSIPYFKWLSFLSWFRYNGALMINQWENVNIQCPNADLPCKK
DGHVILE
TFFHSEADFVMDVVMMLAVLIVGRLVAFLALLVKTWRFK

Advanced options (click to show/hide)

Make Prediction Reset Form

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton G.J. Nucl. Acids Res. (first published online April 16, 2015) doi: 10.1093/nar/gkv332 [link]
More citations: link.

2. 提示 PDB 数据库中存在已知结构，继续预测。



JPred 4
Incorporating Jnet

A Protein Secondary Structure Prediction Server

Home

REST API

About

F.A.Q.

Help & Tutorials

Monitoring

Contact

Publications

Match found in PDB

The sequence you submitted is similar to those with known structure. These may provide a more accurate secondary structure assignment than a JPred prediction.

If you still want to carry out a Jpred prediction click [continue](#)

Hits found

Show **200** entries

PDB	Chain	Description	Blast E-value
7r8c	B	Isoform 4 of ATP-binding cassette	1e-83

3. 预测结果

[illegible]

- 红色的为 α Helix
- 绿色的为 β Sheet
- 其他未无规则卷曲
- α Helix 较多

1. 输入序列后提交

☐ DeepMetaPSICOV 1.0 (Structural Contact Prediction) ☐ MEMPACK (TM Topology and Helix Packing)

Fold Recognition

☐ GenTHREADER (Rapid Fold Recognition) ☐ pDomTHREADER (Protein Domain Fold Recognition)

Structure Modelling

☐ Bioserf 2.0 (Automated Homology Modelling) ☐ Domserf 2.1 (Automated Domain Homology Modelling)

☐ DMPfold 1.0 Fast Mode (Protein Structure Prediction)

Domain Prediction

☐ DomPred (Protein Domain Prediction)

Function Prediction

☐ FFPred 3 (Eukaryotic Function Prediction)

[Help...](#)

Submission details

Protein Sequence

FISKT LAESPIIPIVLTLSVCYFMIGLNSHGFRFYACGIMLVANVAISFGYLSCVSRVSMSLSI
GPPLVIPFLFGGFLNVSIPYFKWLSFLSWFRYNGALMINQWENVNTIQCPNADLPCKPDGHVILE
TFHFSEADFVMDVVMALVIGFRLVAFLLLVKTWRFK

[Help:](#)
If you wish to test these services follow this link to retrieve a test fasta sequence.

Job name

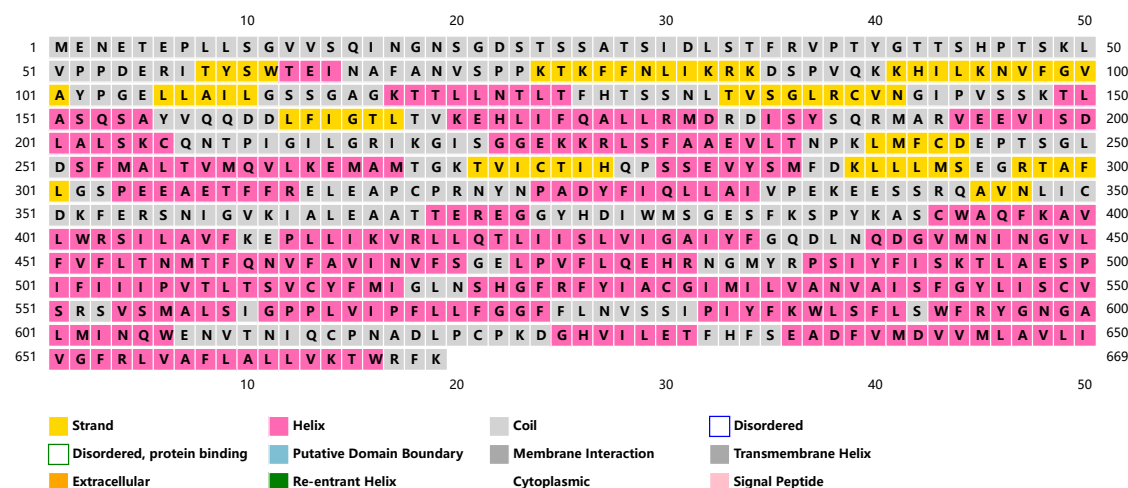
AAL56571.1

Email (optional)

zidongzh@outlook.com

Copyright © 2022.

2. 预测结果



- Helix 较多

PredictProtein [Welcome to PredictProtein!](https://www.predictprotein.org/)

1. 网站下线

PREDICT PROTEIN

Our Apologies...

The PredictProtein service is currently offline due to hardware failure. We're working as quick as we can to resolve the issues.

By our current estimates, PredictProtein will not be back online before January 18th, 2023.

In the meantime, we have a similar and very capable webservice available for you to try - [Lambda Predict Protein \(APP\)](#).

Thanks for your understanding and patience while we resolve this issue.

help@predictprotein.org

使用 Swiss-Model 预测该蛋白质序列的三级结构，并简述预测结果；该蛋白在 AlphaFoldDB 中是否有预测结果？如有，简述预测结果。

1. 输入序列并提交。

The screenshot shows the 'Start a New Modelling Project' interface. It includes a 'Target' section with a list of sequences and their lengths (75, 150, 225, 300, 375). A 'Supported Inputs' dropdown menu is on the right. Below the target list are fields for 'Project Title' (AAL56571.1) and 'Email' (zidongzh@outlook.com). At the bottom, there are 'Search For Templates' and 'Build Model' buttons. A footer section contains a cookie consent message and a timestamp '01:31:22 12/26(1)'.

2. 预测结果

The screenshot shows an email titled 'SWISS-MODEL Modelling Results - AAL56571.1'. The email is from 'noreply-swissmodel@unibas.ch' and is addressed to 'zidongzh@outlook.com'. The body of the email contains the following text:

Hello SWISS-MODEL user,

Modelling results for your modelling project 'AAL56571.1' are now available on the website.

<https://swissmodel.expasy.org/interactive/7usPTm/models>

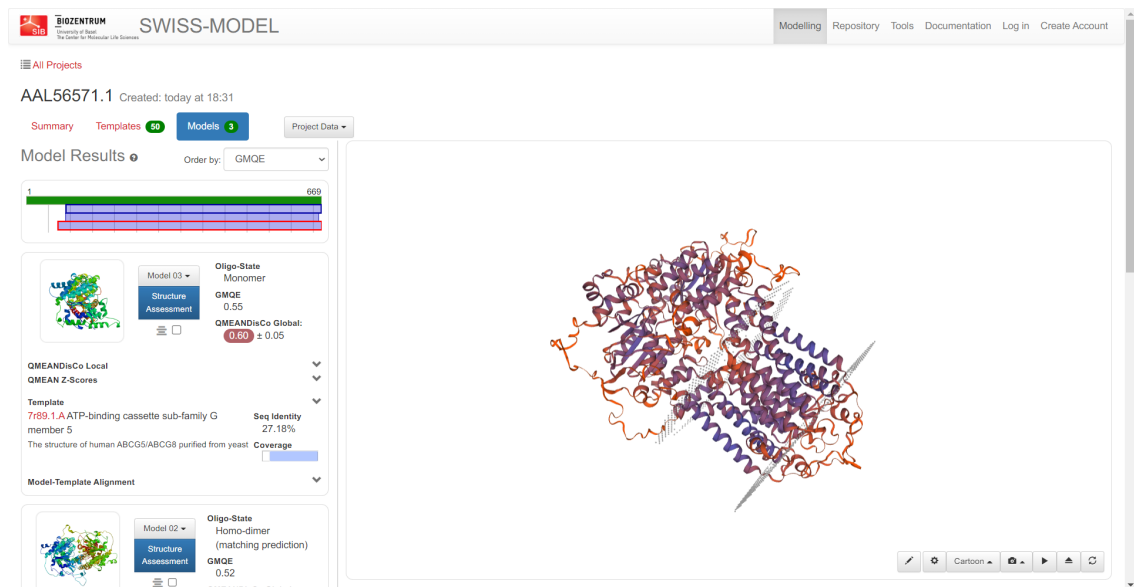
They will remain available for two more weeks until they are deleted. If you would like to have the results available for a longer period, we encourage you to download the results as an archive or request an extension of the project lifetime directly on the project website.

Yours, The SWISS-MODEL team

If you wish to contact us, please send an email to help-swissmodel@unibas.ch

DISCLAIMER:

The result of any theoretical modelling procedure, such as SWISS-MODEL, is NON-EXPERIMENTAL and MUST be considered with care as these models may contain significant errors. This is especially true for automated modelling since there is no human intervention during model building. Please read the header section and the logfile carefully to know what templates and alignments were used during the model building process. All information is provided "AS-IS", without any warranty, expressed or implied.



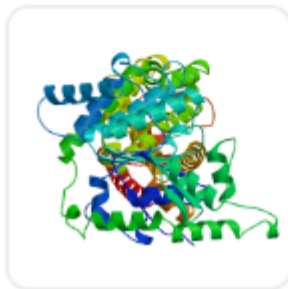
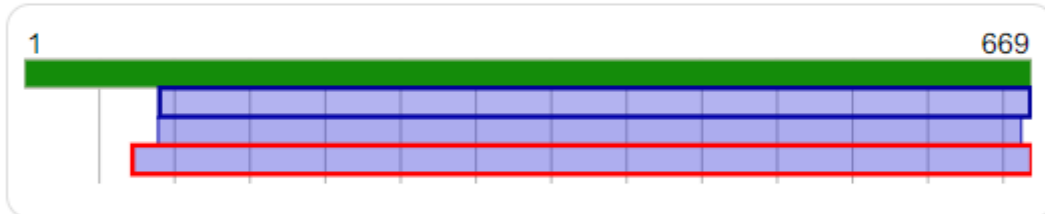
SWISS-MODEL 结果分析

- 模板一致度较低，为 27.18%
- GMQE: 可信度为 0.55

[Summary](#)[Templates](#) **50**[Models](#) **3**[Project Data](#) ▼

Model Results ?

Order by: GMQE ▼



Model 03 ▼

[Structure Assessment](#)**Oligo-State**
Monomer**GMQE**
0.55**QMEANDisCo Global:**
0.60 ± 0.05

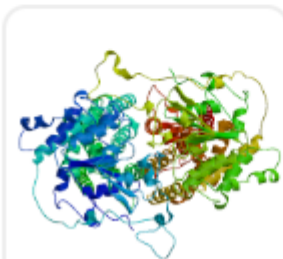
QMEANDisCo Local

QMEAN Z-Scores

Template

7r89.1.A ATP-binding cassette sub-family G member 5

The structure of human ABCG5/ABCG8 purified from yeast

Seq Identity
27.18%**Coverage****Model-Template Alignment**

Model 02 ▼

[Structure Assessment](#)**Oligo-State**
Homo-dimer
(matching prediction)**GMQE**
0.52**QMEANDisCo Global:**

Alpha Fold

在 UniProt 中找到该蛋白 [w - ABC transmembrane transporter white - Tribolium castaneum \(Red flour beetle\)](#) | [UniProtKB](#) | [UniProt](#)

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search

Function
Names & Taxonomy
Subcellular Location
Phenotypes & Variants
PTM/Processing
Expression
Interaction
Structure
Family & Domains
Sequence
Similar Proteins

Low (70 > pLDDT > 50)
Very low (pLDDT < 50)
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
AlphaFold	AF-Q8WRR1-F1	Predicted			1-669	AlphaFold

3D structure databases

Database	Identifier
AlphaFoldDB	Q8WRR1
SMR	Q8WRR1

ModBase Search...

Family & Domains¹

Features

Showing features for region¹, compositional bias¹, domain¹.

1. 在 AlphaFold 中打开

ABC transmembrane transporter white
AlphaFold structure prediction

Download [PDB file](#) [mmCIF file](#) [Predicted aligned error](#)

Note: We have recently updated the PAE JSON format, please refer to our [FAQ](#) for a description of the updated format.

[NEW](#) [Feedback on structure](#) [Looks great](#) [Could be improved](#)

Information

Protein	ABC transmembrane transporter white
Gene	w
Source organism	Tribolium castaneum (Red flour beetle) go to search
UniProt	Q8WRR1 go to UniProt
Experimental structures	None available in the PDB
Biological function	Not available. go to UniProt

3D viewer

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

Sequence of AF-Q8WRR1-F1 Chain 1: ABC transmembrane transporter white

1 11 21 31 41 51 61 71 81 91 101 111
MENETEPFLS GYVSGINGNS GDSTSSATSI DLSTFRVPTT GTTSHTPSKL VPFDERITYS WTEINAFANY SPPTKTFNNL IKRKDSFVQK KHLKNVFGV AYPGELLAIL G
121 131 141 151 161 171 181 191 201 211 221
SSGAGRTTL LNTLTFHTSS NLTVSGLRGV NGIPVSEKTL ASQSAVQDD DLFIGTLTVK EHLIFQALLR NDRDISYSQR MARVEEVIDS LALSKQNTIP IGLGRIGKI SGG
231 241 251 261 271 281 291 301 311 321 331
EKKRLSP AAEVLINPKL NFDCEPTSGI DGFALTVQK VLKEMAMTQK TVICTIHQPS SEVYSMFNNL LLMSEGRATP LGSFPEARTP FPELEACPQR NYNPADYFIQ LLAIV

Alpha Fold 结果分析

- 不同区域的可信度通过不同de

3D viewer

Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Sequence of AF-Q8WRR1-F1 Chain 1: ABC transmembrane transporter white

1 11 21 31 41 51 61 71 81 91 101 111
MENETEPFLS GYVSGINGNS GDSTSSATSI DLSTFRVPTT GTTSHTPSKL VPFDERITYS WTEINAFANY SPPTKTFNNL IKRKDSFVQK KHLKNVFGV AYPGELLAIL G
121 131 141 151 161 171 181 191 201 211 221
SSGAGRTTL LNTLTFHTSS NLTVSGLRGV NGIPVSEKTL ASQSAVQDD DLFIGTLTVK EHLIFQALLR NDRDISYSQR MARVEEVIDS LALSKQNTIP IGLGRIGKI SGG
231 241 251 261 271 281 291 301 311 321 331
EKKRLSP AAEVLINPKL NFDCEPTSGI DGFALTVQK VLKEMAMTQK TVICTIHQPS SEVYSMFNNL LLMSEGRATP LGSFPEARTP FPELEACPQR NYNPADYFIQ LLAIV

讨论

Nanog

- 亚细胞定位
 - Deeploc 结果最清晰明了。
- 是否糖蛋白
 - GlycoEP 相比其他两种方法运行较慢。提交此实验报告时未受到结果邮件。

分析甲虫基因 AF422804

- 二级结构预测
 - PSIRPRED 预测结果更直观。
 - PredictProtein 网站下线
- 三级结构预测
 - 不知道如何在 Alpha Fold DB 中直接搜索该蛋白，最后通过 UniProt 搜索转到 Alpha Fold 中。