



中国科学院  
CHINESE ACADEMY OF AGRICULTURAL SCIENCES

# 黑麦草抗冻蛋白的生物信息学分析

Bioinformatics analysis of antifreeze proteins in ryegrass

- 汇报人：苏德毕力格
- 小组：2G08
- 2022年5月 14日

明德格物 博学笃行



①

背景

②

系统进化树比较

③

基因功能分析

④

蛋白结构分析



## 黑麦草

**黑麦草**是禾本科黑麦草属的多年生植物，其具有营养高、生长快、分蘖多、能耐牧的特性，是世界各地普遍引种栽培的优良牧草。

黑麦草的生长周期快，而且草茎纤细柔嫩，具有较好的耐践踏性，也是高尔夫球场的常用草。



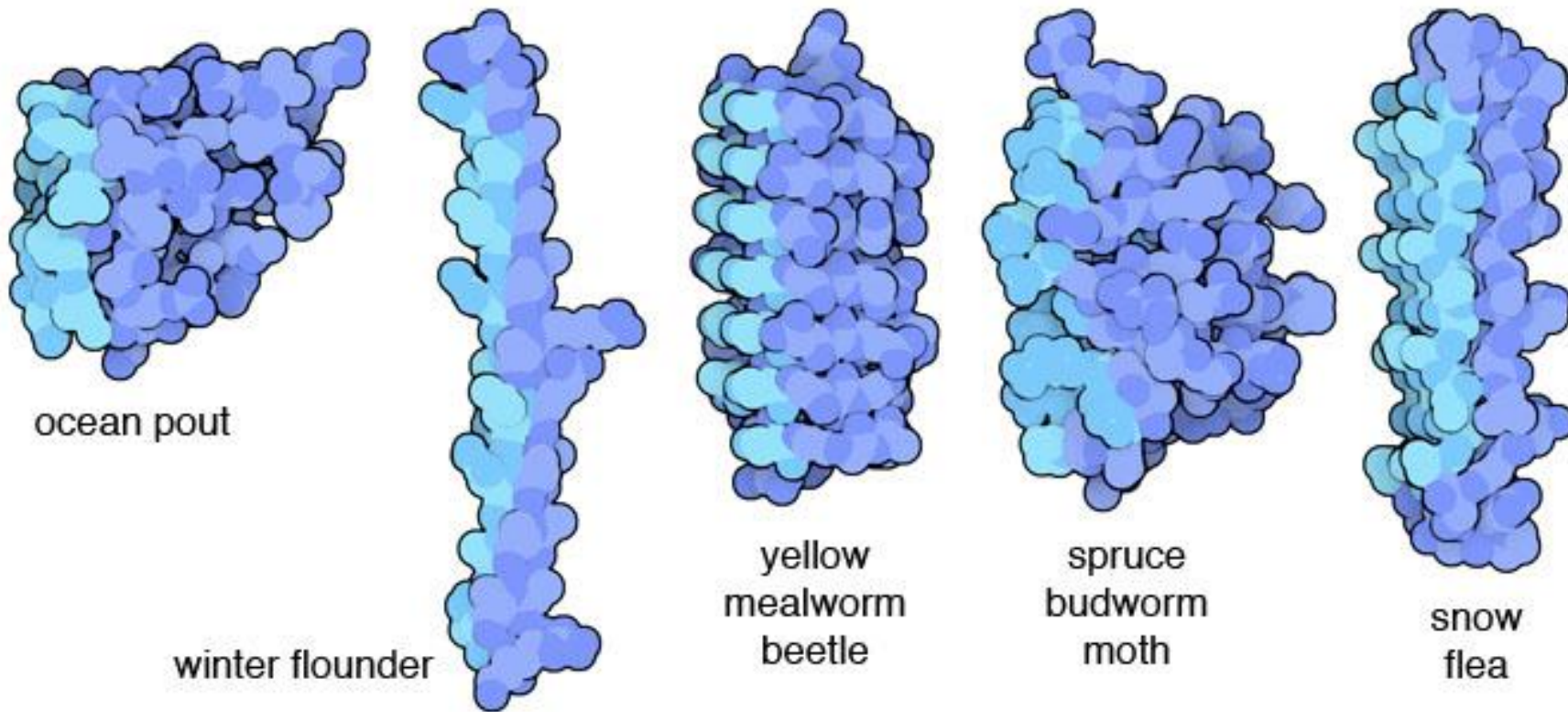
## 背景——抗冻蛋白的发现

**抗冻蛋白**最初是从南极与北极地区的海洋鱼类血清中发现的，它能阻止体液内冰核的形成与生长，维持体液的非冰冻状态，以适应低温的生活条件。现在相继在昆虫、植物(如冬黑麦、胡萝卜、小麦等)体内也发现有类似功能的抗冻蛋白。

从北极鱼抗冻蛋白(Antifreeze proteins, AFPs)基因(AFP)被克隆以来,国内外学者曾尝试用转鱼类AFP基因的方法来提提高作物抗冻性。由于鱼类和植物在分类学上的关系相差太远,外源基因的正确表达有一定困难,研究结果不甚理想。在此情况下,只得寄希望于植物AFP基因的发现。



# 背景——抗冻蛋白的发现



抗冻蛋白是趋同进化的完美例子，他们发挥相似的功能。这些表面平坦的蛋白质，富含苏氨酸（浅蓝色），可与冰晶表面结合。



## 研究目的

**植物抗冻蛋白**是从许多抗冻植物中分离的、参与植物抵御冻害反应的一类新型蛋白。这类抗冻蛋白具有多个亲水性缚冰域,能直接作用于冰晶,阻止冰晶在细胞间隙形成和再结晶。

国内外对植物抗冻蛋白的研究开展得比较晚,但进展却很迅速,相关方面在已研究或正研究的植物中,以冬黑麦抗冻蛋白研究得最为详细。

1992年首次在低温锻炼的**冬黑麦(Secale cereale)**中发现内源性抗冻蛋白。

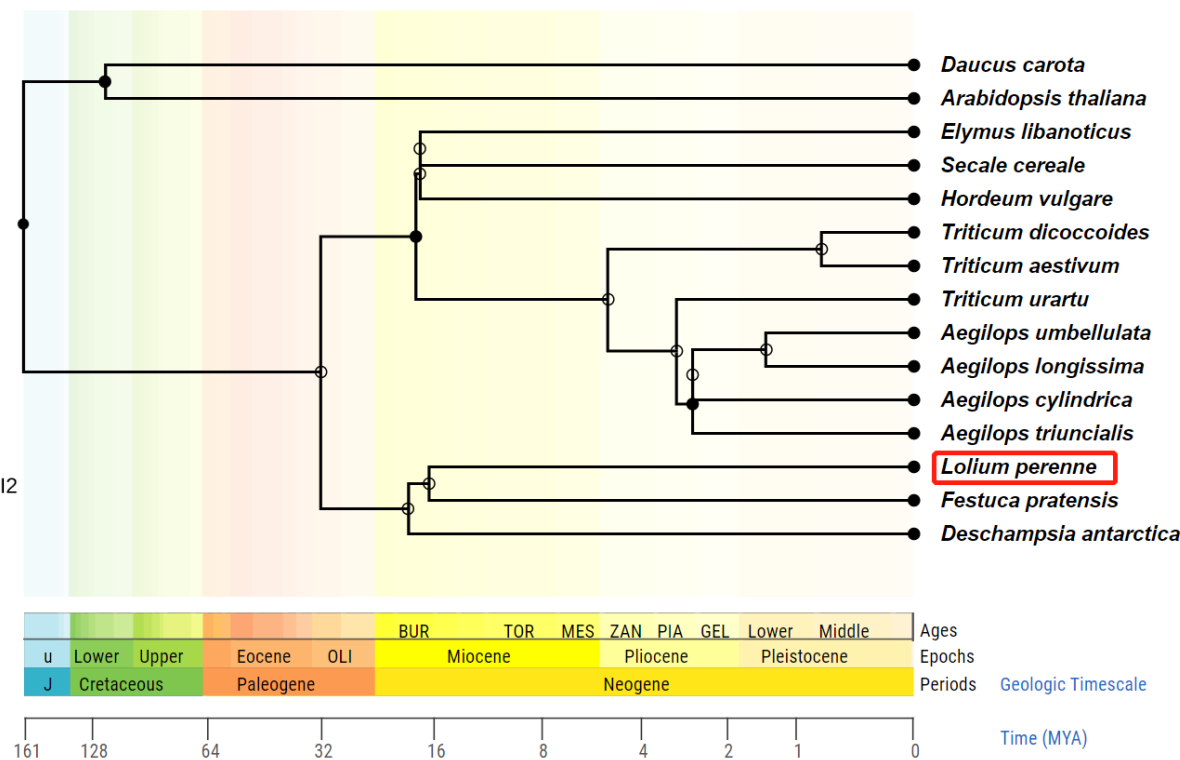
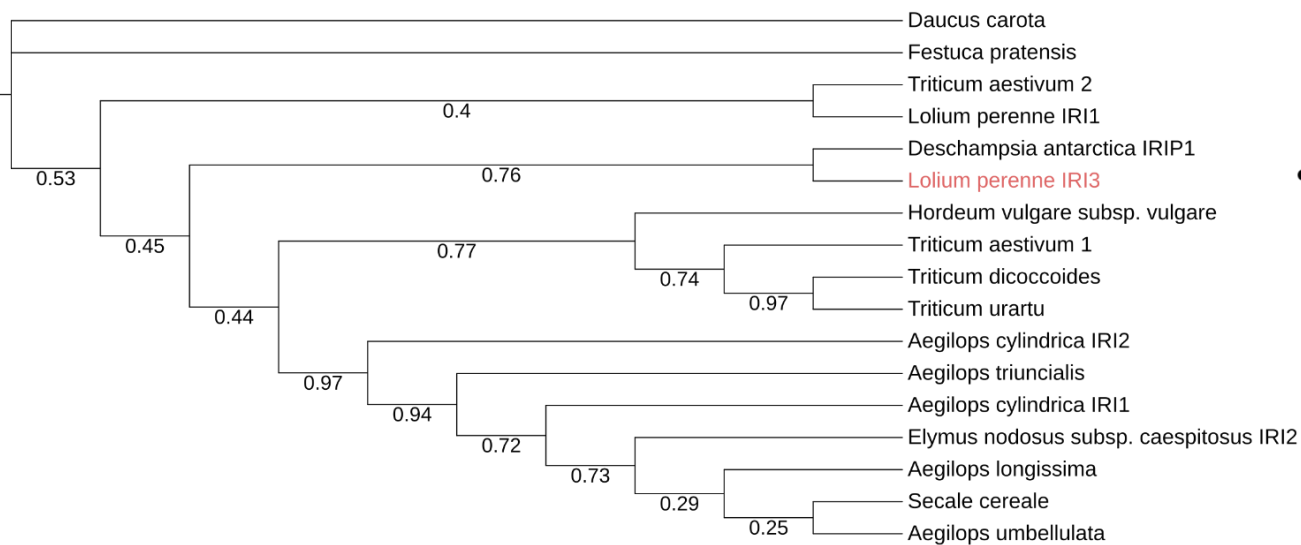


# 抗冻蛋白基因信息

物种	蛋白质中文名 (英文名)	序列条目名	基因名	长度
<b>黑麦草</b> <b>Lolium perenne</b> (Perennial ryegrass)	冰重结晶抑制蛋白 Ice recrystallization inhibition protein-like protein	B5T007_LOLPR	IRI3	254
<b>胡萝卜</b> <b>Daucus carota</b> (Wild carrot)	抗冻蛋白 AFP protein	O82438_DAUCA	AFP	332
<b>小麦</b> <b>Triticum aestivum</b> (Wheat)	冰重结晶抑制蛋白2 Ice recrystallization inhibition protein 2	Q56B89_WHEAT	N/A	409
	冰重结晶抑制蛋白1 Ice recrystallization inhibition protein 1	Q56B90_WHEAT	N/A	280



# 系统进化树比较



抗冻蛋白系统进化树

Timetree物种的时间分化

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





# 基因分析

Name	p-value	Motif Locations
tr Q56B90 Q56B90_WHEAT	3.45e-110	
tr Q56B89 Q56B89_WHEAT	5.15e-104	
tr O82438 O82438_DAUCA	1.51e-31	
tr B5T007 B5T007_LOLPR	2.67e-105	

Motif	Symbol	Motif Consensus
1.		WSGASCCDWEGV
2.		MAFTNMPLYVKR
3.		GPIPSWIGZLDH



E-value	Sites	Width	M
1.2e-024	4	27	
6.7e-027	3	49	
1.9e-020	4	29	

<https://meme-suite.org/meme/>



# 基因分析

tr|O82438|O82438\_DAUCA M N I E S S F C P I L C I C M I F L C L P - N L S A S Q R C N N N D K Q A L L Q I K T A L K N - - - P T I T D S W V S D D 57  
 tr|Q56B89|Q56B89\_WHEAT - - - - M A K S W L L L F C L A F L L P A A - - - - S A T C H P D D L R A L R G L A G T L G A V A - R H L H S V W - S G A 51  
 tr|Q56B90|Q56B90\_WHEAT - - - - M A K C G L L L L F L A F L L P A A - - - R A T S C H P D D L R A L R G F A G N L S G G A A - L L R A A W - S G A 52  
 tr|B5T007|B5T007\_LOLPR - - - - M A K C L M L L L S F A F L L S A A G T A T A T P C H R D D L R A L R G F A E N L G G G G A L S L R A A W - S G A 56

B5T007:Domain

tr|O82438|O82438\_DAUCA D C C G W D L V E C D E T S N R I I S L I I Q D D E A L T G Q I P P Q - V A L W F R K L P N L F G K I P E E 117  
 tr|Q56B89|Q56B89\_WHEAT S C C D W E G V G C H G A S G R V T A L R L H G H - G L A G P I P G A S L A G L I R L E E L F - L G S N S F A G T L P D E 110  
 tr|Q56B90|Q56B90\_WHEAT S C C V W E G V N C D G T S G R V T A L R L P G H - G L V G L I P G A S L L E E L - - - - - - - - - - 97  
 tr|B5T007|B5T007\_LOLPR S C C D W E G V G C D G A S G R V T A L W L P R S - G L - - - - - - - - - - - - - - - - 83


B5T007:Domain


**Domain 29-67**

Description

LRRNT\_2

<https://beta.uniprot.org/align>

1 - 1 of 1 entry matching Pfam 

ACCESSION	NAME	SOURCE DATABASE	MATCHES
PF08263	LRRNT_2	Pfam	

富含亮氨酸重复序列(LRR)是短序列基序，存在于许多具有不同功能和细胞位置的蛋白质中。富含亮氨酸的重复序列通常侧接富含半胱氨酸的结构域。该结构域通常位于富含亮氨酸的串联重复序列的 N 末端。与某些蛋白质互作相关。

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



# 基因分析

将黑麦草抗冻蛋白序列在ref-protein数据库中进行Blast比对，结果如下：

Sequences producing significant alignments

Download

Select columns

Show

100



select all 100 sequences selected

[GenPept](#)

[Graphics](#)



[Distance tree of results](#)

[Multiple alignment](#)

[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">phytosylfokine-alpha receptor 2 [Arabidopsis thaliana]</a>	<a href="#">Arabidopsis thali...</a>	72.0	72.0	51%	3e-13	39.10%	1036	<a href="#">NP_200200.1</a>
<input checked="" type="checkbox"/>	<a href="#">LRR receptor-like serine/threonine-protein kinase FEI 1 precursor [Zea mays]</a>	<a href="#">Zea mays</a>	56.6	56.6	34%	4e-08	37.00%	594	<a href="#">NP_001146833.1</a>
<input checked="" type="checkbox"/>	<a href="#">LRR receptor-like serine/threonine-protein kinase FEI 1 precursor [Zea mays]</a>	<a href="#">Zea mays</a>	55.5	55.5	34%	9e-08	37.00%	594	<a href="#">NP_001147917.2</a>
<input checked="" type="checkbox"/>	<a href="#">receptor-like kinase [Arabidopsis thaliana]</a>	<a href="#">Arabidopsis thali...</a>	55.5	55.5	34%	1e-07	42.05%	751	<a href="#">NP_201529.1</a>
<input checked="" type="checkbox"/>	<a href="#">Probable LRR receptor-like serine/threonine-protein kinase At4g37250 [Zea mays]</a>	<a href="#">Zea mays</a>	55.1	97.0	34%	1e-07	38.00%	819	<a href="#">NP_001169585.1</a>
<input checked="" type="checkbox"/>	<a href="#">Leucine-rich repeat protein kinase family protein [Arabidopsis thaliana]</a>	<a href="#">Arabidopsis thali...</a>	53.9	53.9	22%	3e-07	42.68%	627	<a href="#">NP_186938.1</a>

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

IPR032675	LRR_dom_sf	InterPro		100
IPR013210	LRR_N_plant-typ	InterPro		100



# 基因分析

Number of amino acids: 254

Molecular weight: 26248.00

Theoretical pI: 7.11

Amino acid composition:

CSV

Ala (A)	17	6.7%
Arg (R)	11	4.3%
Asn (N)	31	12.2%
Asp (D)	12	4.7%
Cys (C)	6	2.4%
Gln (Q)	4	1.6%
Glu (E)	4	1.6%
Gly (G)	31	12.2%
His (H)	9	3.5%
Ile (I)	6	2.4%
Leu (L)	24	9.4%
Lys (K)	5	2.0%
Met (M)	4	1.6%
Phe (F)	4	1.6%
Pro (P)	7	2.8%
Ser (S)	31	12.2%
Thr (T)	18	7.1%
Trp (W)	4	1.6%
Tyr (Y)	1	0.4%
Val (V)	25	9.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

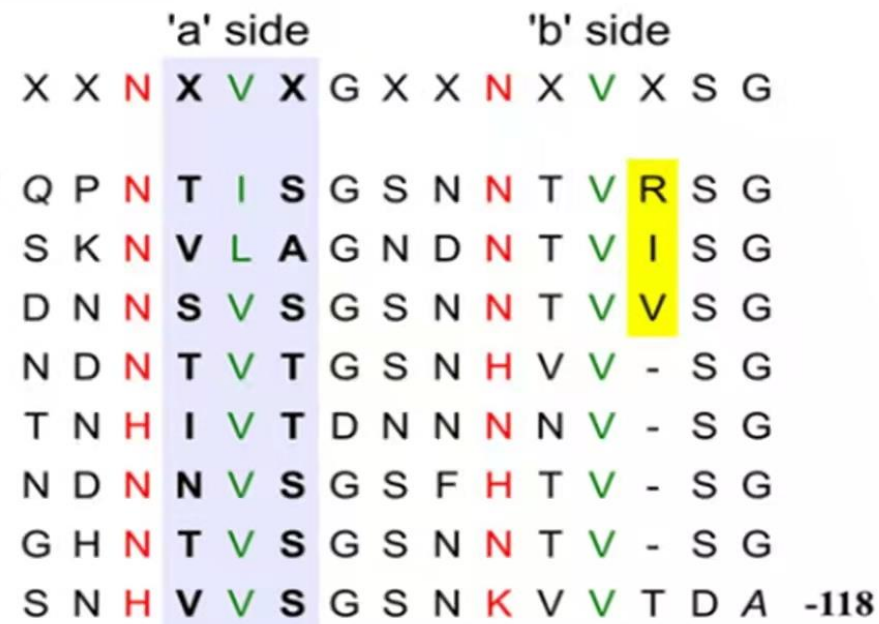
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

## 6.1 Composition in percent for the complete database

Ala (A)	8.25	Gln (Q)	3.93	Leu (L)	9.65	Ser (S)	6.64
Arg (R)	5.53	Glu (E)	6.72	Lys (K)	5.80	Thr (T)	5.35
Asn (N)	4.06	Gly (G)	7.07	Met (M)	2.41	Trp (W)	1.10
Asp (D)	5.46	His (H)	2.27	Phe (F)	3.86	Tyr (Y)	2.92
Cys (C)	1.38	Ile (I)	5.91	Pro (P)	4.74	Val (V)	6.86

<https://web.expasy.org/docs/relnotes/re.html>

该蛋白富含天冬酰胺 (12.2%)、缬氨酸 (9.8%)、丝氨酸 (12.2%)、苏氨酸 (7.1%)，而很少带芳香基团的氨基酸，表明黑麦草抗冻蛋白的亲水性较强。预测有两个冰晶结合面a面和b面，每面的模体“XXNXVXG”与亲疏水性有关，都以缬氨酸 (V) 为冰晶结合中心。



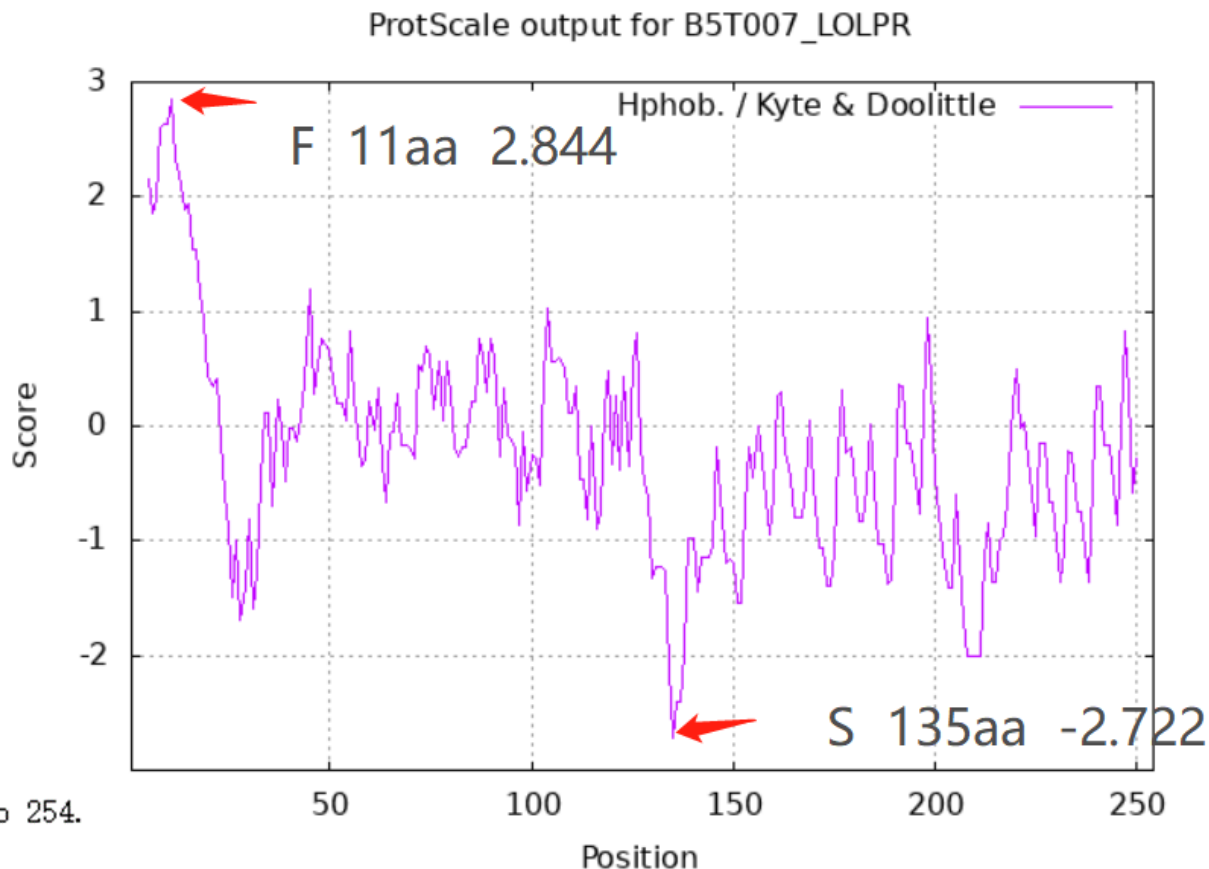
(Adam J. Middleton, 2012)

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



# 蛋白结构分析——二级结构分析

**亲疏水性预测分析：**  
氨基酸表现为亲水性时数值为负值，表现为疏水性时数值为正，数值绝对值越高表明亲疏水性越强。



MIN: -2.722, MAX: 2.844  
Sequence: B5T007 (B5T007\_LOLPR) from 1 to 254.  
Scale: Hydropathicity.

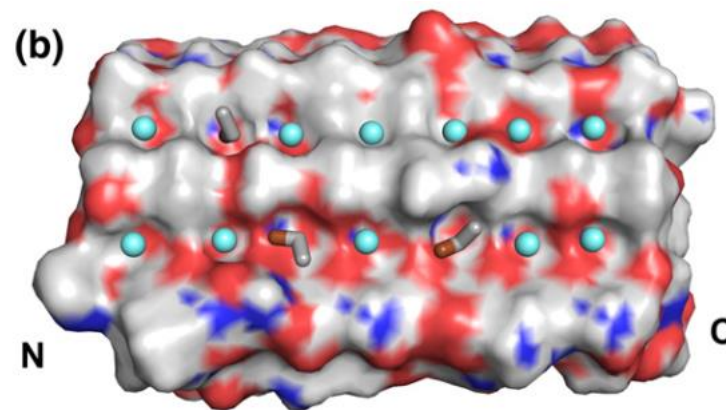
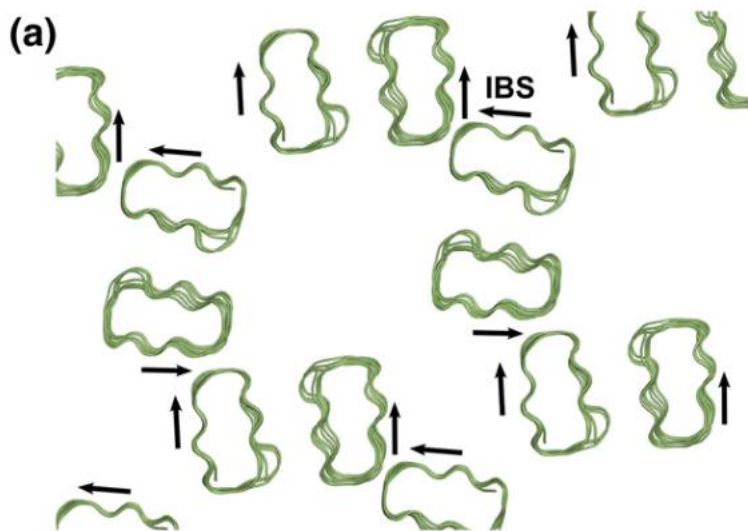
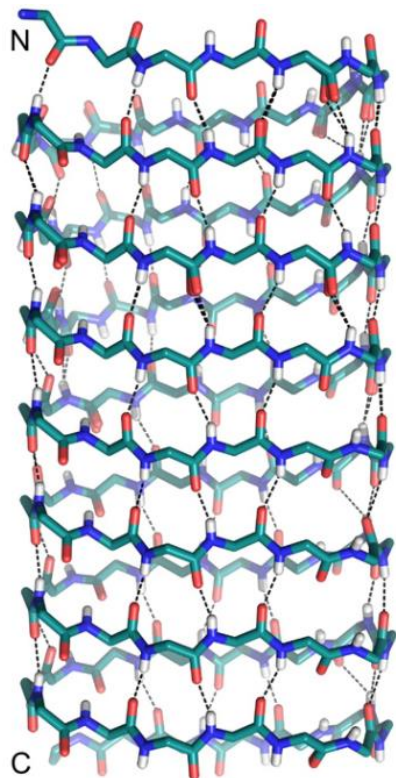
Window size: 9  
Scale not normalized.  
Relative weight for window edges: 100 %  
Weight variation model: linear

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



# 蛋白结构分析——三级结构分析

碳原子  
氮原子  
氧原子



白色的是碳，蓝色的是氮，红色的是氧。浅蓝色的球体显示了LpIBP结构的IBS (ice binding surface) 上有序水分子的位置。灰色和棕色的棍子显示了在IBS上发现的乙醇分子的位置。

LpIBP主要通过外部和内部Asn/His阶梯内的广泛氢键网络以及一个小的疏水核心来稳定。

(Adam J. Middleton, 2012)



中国农业科学院  
CHINESE ACADEMY OF AGRICULTURAL SCIENCES

明德格物 博学笃行

THANK YOU FOR YOUR CRITICISM  
AND CORRECTION