

拟南芥中小分子多肽IDL5 的序列、功能和结构分析

2G03: 邵煜尊 王红 薛瑾 张哲

2022年5月14日



目录

CONTENTS

01

研究背景

02

序列、功能分析

03

理化性质与结构预测

01 Background

- 1、**多肽**是动植物体内的一类具有**调控**作用的**小分子物质**
- 2、由前体蛋白经过翻译后剪切、修饰，最终形成只有**较短氨基酸**组成的**成熟多肽**，作为配体与定位在细胞膜上的受体识别，启动信号转导，进而调控多种生物学过程。
- 3、**IDA家族**的小分子多肽能够与类受体酶结合，通过影响细胞壁重建相关酶的表达，引起细胞分离，**调控花器官脱落过程**。

01 Background

IDL: IDA-LIKE

IDA:

英文全称: inflorescence deficient in abscission

中文含义: 花序脱落缺陷

功能: 拟南芥中IDA突变, 阻碍了花器官的脱落

它是通过富含亮氨酸的重复受体激酶(LRR-RLKs)

HAESA (HAE)和HAESA-like2 (HSL2)信号来调控授粉后花的脱落

01 Background

因此，利用生信技术，了解IDL家族中**IDL5**小分子多肽的相关信息，为下一步实验做铺垫

(一) Analyzed by NCBI

对IDL5进行核苷酸条目搜索

- [Arabidopsis thaliana chromosome 1, whole genome shotgun sequence](#)
1. 29,387,870 bp linear DNA
Accession: LUHQ01000001.1 GI: 1032297437
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- [Arabidopsis thaliana inflorescence deficient in abscission \(IDA\)-like 5 \(IDL5\), mRNA](#)
2. 460 bp linear mRNA
Accession: NM_001334749.1 GI: 1063678629
[BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- [Arabidopsis thaliana ida like-protein 5 \(IDL5\) gene, complete cds](#)
3. 312 bp linear DNA
Accession: AY642386.1 GI: 49425315
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- [Arabidopsis thaliana chromosome 1 sequence](#)
4. 30,427,671 bp linear DNA
Accession: NC_003070.9 GI: 240254421
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- [Arabidopsis thaliana chromosome 1 sequence](#)
5. 30,427,671 bp linear DNA
Accession: CP002684.1 GI: 332189094
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- [Dolichospermum lemmermannii FEM IDL5 partial rpoB gene for RNA polymerase B subunit](#)
6. 575 bp linear DNA
Accession: LN871477.1 GI: 952007805
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

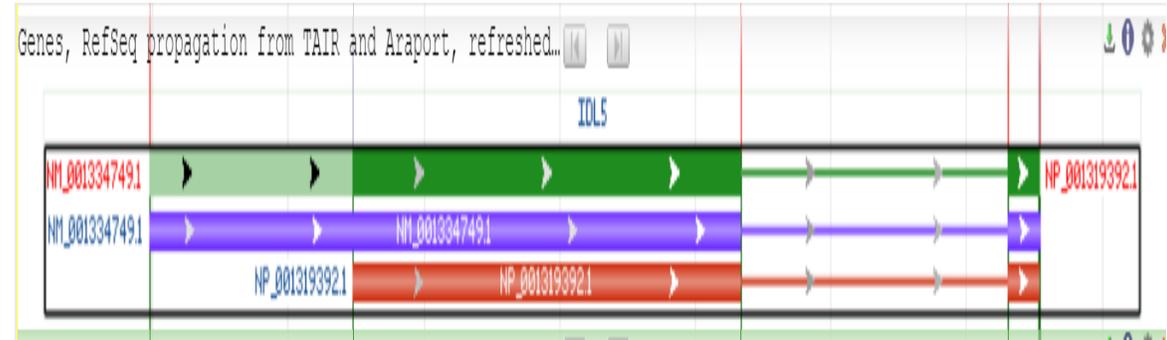
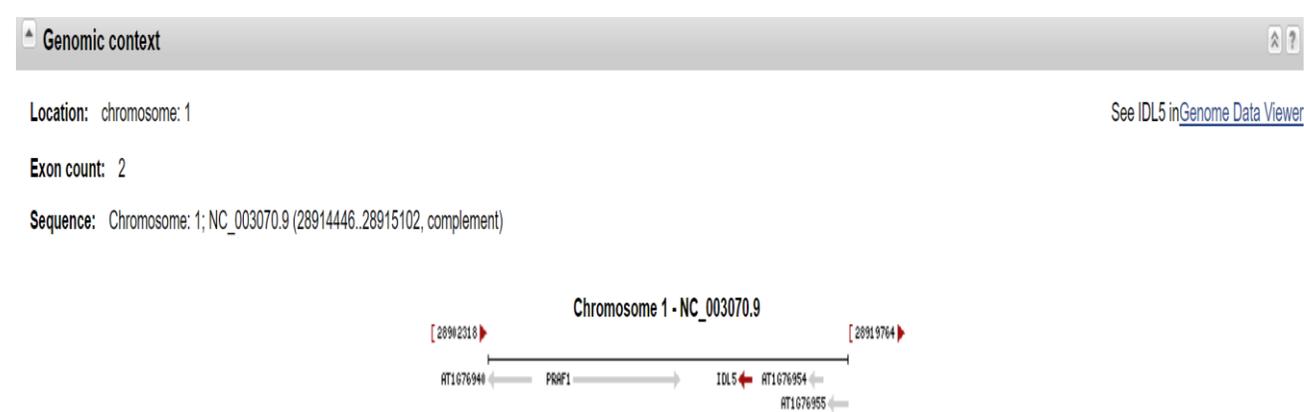
- [Dolichospermum lemmermannii FEM IDL5 partial 16S rRNA gene](#)
7. 1,454 bp linear DNA
Accession: LN871463.1 GI: 952007784
[Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- [Arabidopsis thaliana ecotype Landsberg erecta chromosome 1, whole genome shotgun sequenc](#)
8. 29,387,870 bp linear DNA
Accession: CM004359.1 GI: 1032966058
[Assembly](#) [BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)

搜索到8条有关IDL5的信息，其中有6条是跟拟南芥有关，剩下2条是跟细菌有关

(一) Analyzed by NCBI

了解拟南芥中IDL5基因的相关信息



基因组结构：位于1号染色体上，有两个外显子

IDL5基因序列长度为657nt；转录mRNA序列长度为460nt；

编码区序列长度为309nt

(二) Analyzed by uniprot

高级检索拟南芥中IDL家族的其他基因

基因名	蛋白名	长度	功能
IDL1	Protein IDL-LIKE 1	86	Involved in an ethylene-independent separation step of floral abscission. May act with RLK5 and HSL2 as ligand-receptor pairs
IDL2	Protein IDL-LIKE 2	95	May be involved in floral abscission
IDL3	Protein IDL-LIKE 3	99	May be involved in floral abscission
IDL4	Protein IDL-LIKE 4	93	May be involved in floral abscission.
IDL5	Protein IDL-LIKE 5	111	May be involved in floral abscission.

(二) Analyzed by uniprot

了解拟南芥中Protein IDL-LIKE 5的相关信息

Protein IDL-LIKE 5	注释
功能	Similar to Inflorescence deficient in abscission (IDA). Involved in floral organ abscission
亚细胞定位	Extracellular region or secreted
信号肽	Position1-27
组织特异性	Expressed mainly in flowers. Lower levels in buds and seedlings (芽和幼苗) .Detected in vascular tissues and in hydathodes (维管组织和水螅)
相互作用	与IDL1、IDL2、IDL4、 CEP15 等有蛋白互作
基因序列	有1个可能的亚型映射: Inflorescence deficient in abscission (IDA)-like 5
相似蛋白	ARALYDRAFT_316610 、IDL5、 Inflorescence deficient in abscission (IDA)-like 5

(三) Analyzed by BLASTP

利用拟南芥IDL5基因序列进行blastp搜索

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments

Download | Select columns | Show 100

select all 51 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
RecName: Full=Protein IDA-LIKE 5; Flags: Precursor [Arabidopsis thaliana]	Arabidopsis thaliana	143	143	100%	2e-44	100.00%	111	Q6DUW8.2

得到了51条相关的序列信息

Sequences producing significant alignments

Download | Select columns | Show 1000

select all 3 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"/> inflorescence deficient in abscission (IDA)-like 5 [Arabidopsis thaliana]	Arabidopsis thaliana	131	131	100%	2e-39	100.00%	102	NP_001319392.1
<input checked="" type="checkbox"/> RecName: Full=Protein IDA-LIKE 5; Flags: Precursor [Arabidopsis thaliana]	Arabidopsis thaliana	121	121	96%	2e-35	98.98%	111	Q6DUW8.2
<input checked="" type="checkbox"/> ida like-protein 5 [Arabidopsis thaliana]	Arabidopsis thaliana	119	119	96%	9e-35	97.96%	103	AAT66016.1

选择拟南芥，
得到3条相关信息，
进行alignment

(三) Analyzed by BLASTP

Alignment结果的分析

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

inflorescence deficient in abscission (IDA)-like 5 [Arabidopsis thaliana]

Sequence ID: [NP_001319392.1](#) Length: 102 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 102 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
131 bits(479)	2e-39	Compositional matrix adjust.	102/102(100%)	102/102(100%)	0/102(0%)
Query 1	MGNKRIKAmmlilvmmimvFSWRICEADSLRRYS	SSSRPQRF	FKVRRPNPRNH	HHQNGQF	60
Sbjct 1	MGNKRIKAMMILVVMIMMVFSWRICEADSLRRYS	SSSRPQRF	FKVRRPNPRNH	HHQNGQF	60
Query 61	NGDDYPPESFSGFLPKTLP	IPHSAPSRKH	NVYGLQR	INYSCH	102
Sbjct 61	NGDDYPPESFSGFLPKTLP	IPHSAPSRKH	NVYGLQR	INYSCH	102

Related Information

[Gene](#) - associated gene details
[Genome Data Viewer](#) - aligned genomic context
[Identical Proteins](#) - Identical proteins to NP_001319392.1

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

RecName: Full=Protein IDA-LIKE 5; Flags: Precursor [Arabidopsis thaliana]

Sequence ID: [Q6DUW8.2](#) Length: 111 Number of Matches: 1

Range 1: 1 to 98 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
121 bits(442)	2e-35	Compositional matrix adjust.	97/98(99%)	97/98(98%)	0/98(0%)
Query 1	MGNKRIKAmmlilvmmimvFSWRICEADSLRRYS	SSSRPQRF	FKVRRPNPRNH	HHQNGQF	60
Sbjct 1	MGNKRIKAMMILVVMIMMVFSWRICEADSLRRYS	SSSRPQRF	FKVRRPNPRNH	HHQNGQF	60
Query 61	NGDDYPPESFSGFLPKTLP	IPHSAPSRKH	NVYGLQR	IN	98
Sbjct 61	NGDDYPPESFSGFLPKTLP	IPHSAPSRKH	NVYGLQR	N	98

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

ida like-protein 5 [Arabidopsis thaliana]

Sequence ID: [AAT66016.1](#) Length: 103 Number of Matches: 1

[See 2 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 98 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
119 bits(435)	9e-35	Compositional matrix adjust.	96/98(98%)	96/98(97%)	0/98(0%)
Query 1	MGNKRIKAmmlilvmmimvFSWRICEADSLRRYS	SSSRPQRF	FKVRRPNPRNH	HHQNGQF	60
Sbjct 1	MGNKRIKAMMILVVMIMMVFSWRICEADSLRRYS	SSSRPQRF	FKVRRPNPRNH	HHQNGQF	60
Query 61	NGDDYPPESFSGFLPKTLP	IPHSAPSRKH	NVYGLQR	IN	98
Sbjct 61	NGDDYPPESFSGFLPKTLP	IPHSAPSRKH	NVYGLQ	STN	98

Related Information

[Identical Proteins](#) - Identical proteins to AAT66016.1

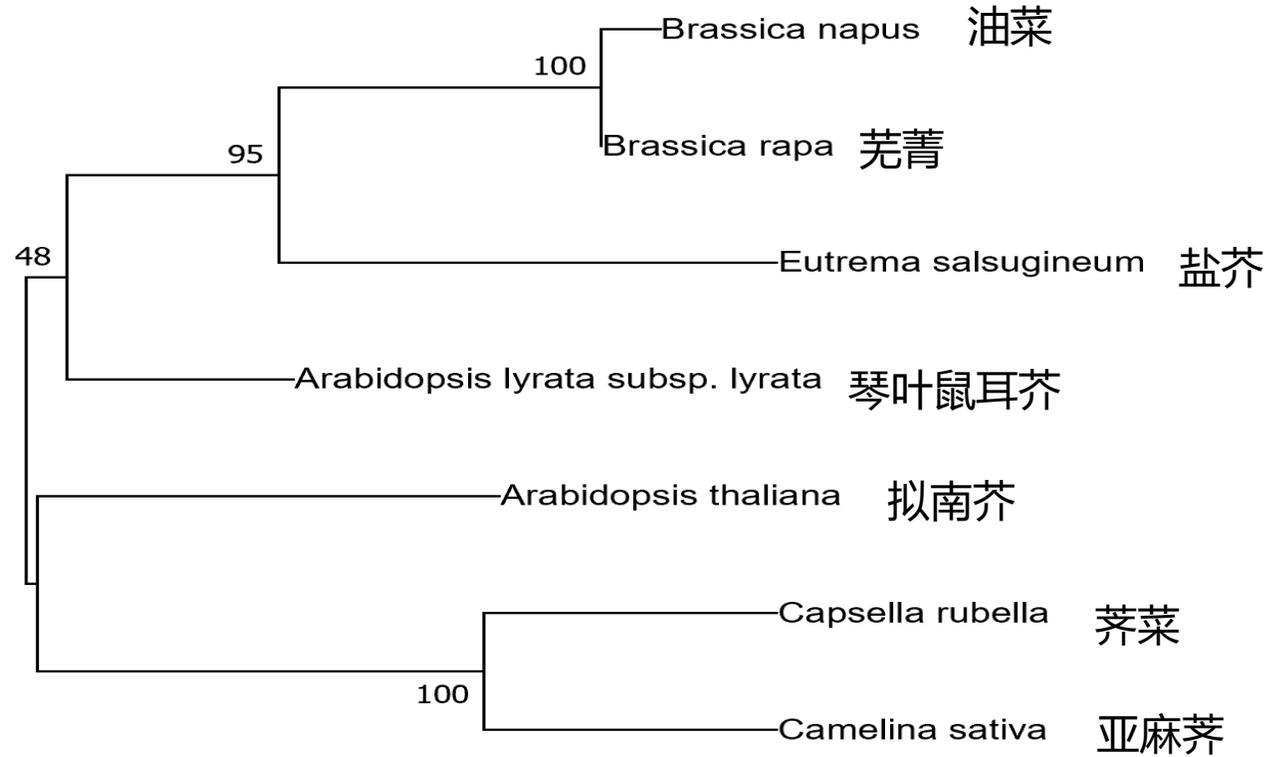
(四) Analyzed by MAGE

IDL5在不同物种中的相关信息

物种名	拉丁名	Accession	AA长度	Gene ID
拟南芥	<i>Arabidopsis thaliana</i>	Q6DUW8.2	111	12972671
亚麻荠	<i>Camelina sativa</i>	XP_010471861.1	98	104751588
荠菜	<i>Capsella rubella</i>	XP_023644012.1	101	17896165
盐芥	<i>Eutrema salsugineum</i>	XP_024012523.1	99	18008838
芜菁	<i>Brassica rapa</i>	XP_033132028.1	102	103830585
油菜	<i>Brassica napus</i>	XP_013713292.1	102	106416963
琴叶鼠耳芥	<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>	XP_020891483.1	101	9323728

(四) Analyzed by MAGE

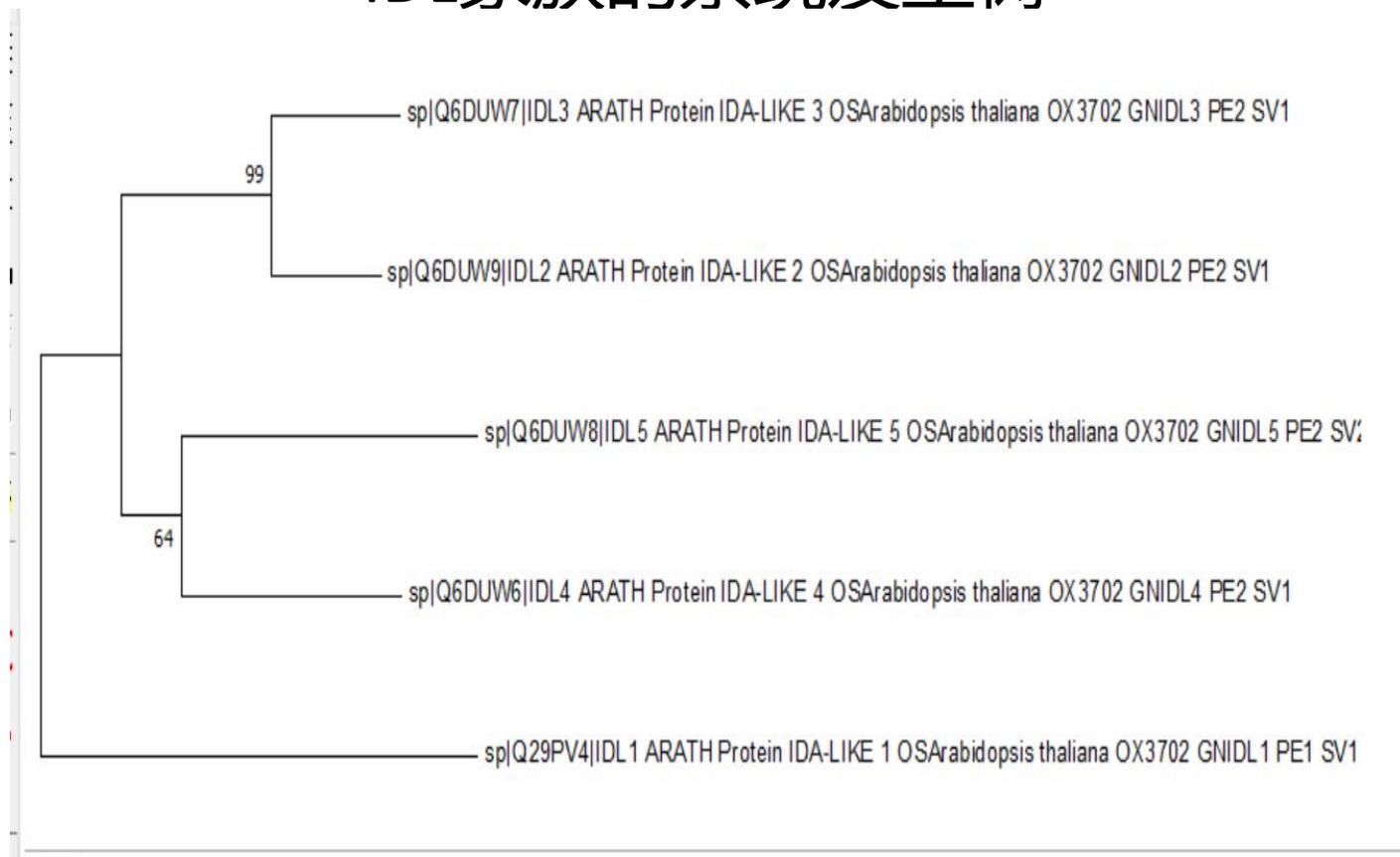
IDL5在不同物种的系统发生树



└───┬───┘ 油菜和芜菁的亲缘关系较近，亚麻芥和荠菜的亲缘关系较近

(四) Analyzed by MAGE

IDL家族的系统发生树



IDL2和IDL3序列相似性更高一些、 IDL4和IDL5序列相似性更高一些

(一) 理化性质预测分析

用Exrasy在线网站 (<https://web.expasy.org/protparam/>) 对其进行理化性质分析

Number of amino acids: 111

Molecular weight: 13059.98

Theoretical pI: 11.18

Amino acid composition:

CSV format

Ala (A)	4	3.6%
Arg (R)	12	10.8%
Asn (N)	7	6.3%
Asp (D)	3	2.7%
Cys (C)	1	0.9%
Gln (Q)	4	3.6%
Glu (E)	2	1.8%
Gly (G)	5	4.5%
His (H)	7	6.3%
Ile (I)	5	4.5%
Leu (L)	6	5.4%
Lys (K)	5	4.5%
Met (M)	6	5.4%
Phe (F)	7	6.3%
Pro (P)	9	8.1%
Ser (S)	14	12.6%
Thr (T)	3	2.7%
Trp (W)	2	1.8%
Tyr (Y)	4	3.6%
Val (V)	5	4.5%
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): 5

Total number of positively charged residues (Arg + Lys): 17

Instability index:

The instability index (II) is computed to be 57.33

This classifies the protein as unstable.

氨基酸数量111个；分子量为13059.98；理论等电点为11.18

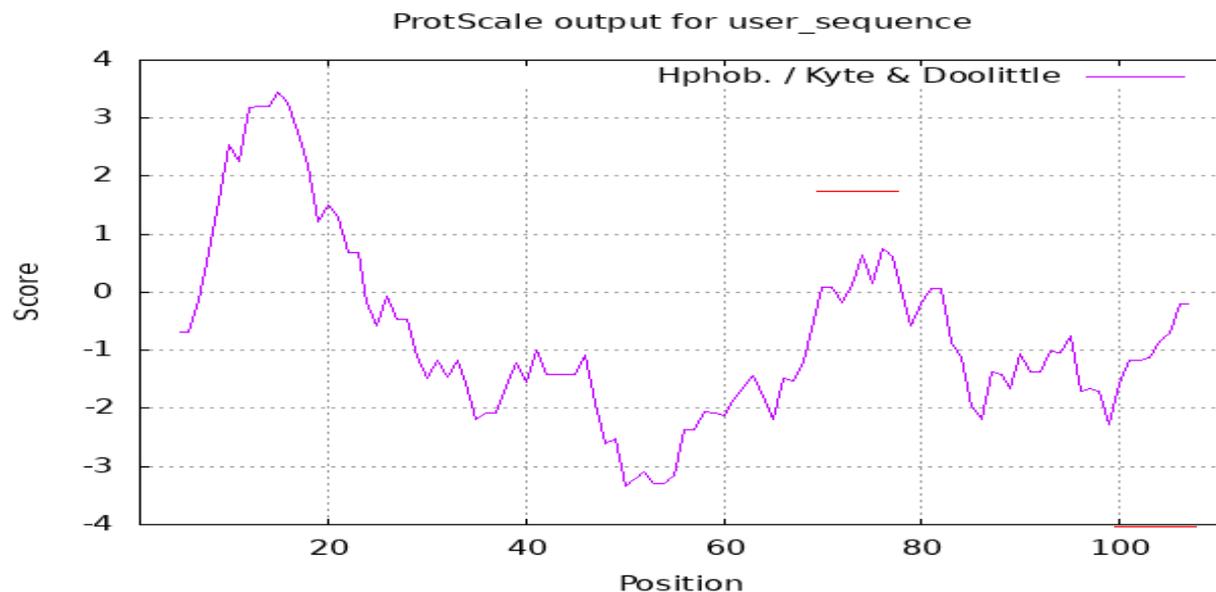
带负电残基 (Asp+Glu) 数量: 5个

带正电残基 (Arg+Lys) 数量: 17个

稳定系数: 57.33 (该蛋白不稳定)

理化性质预测分析

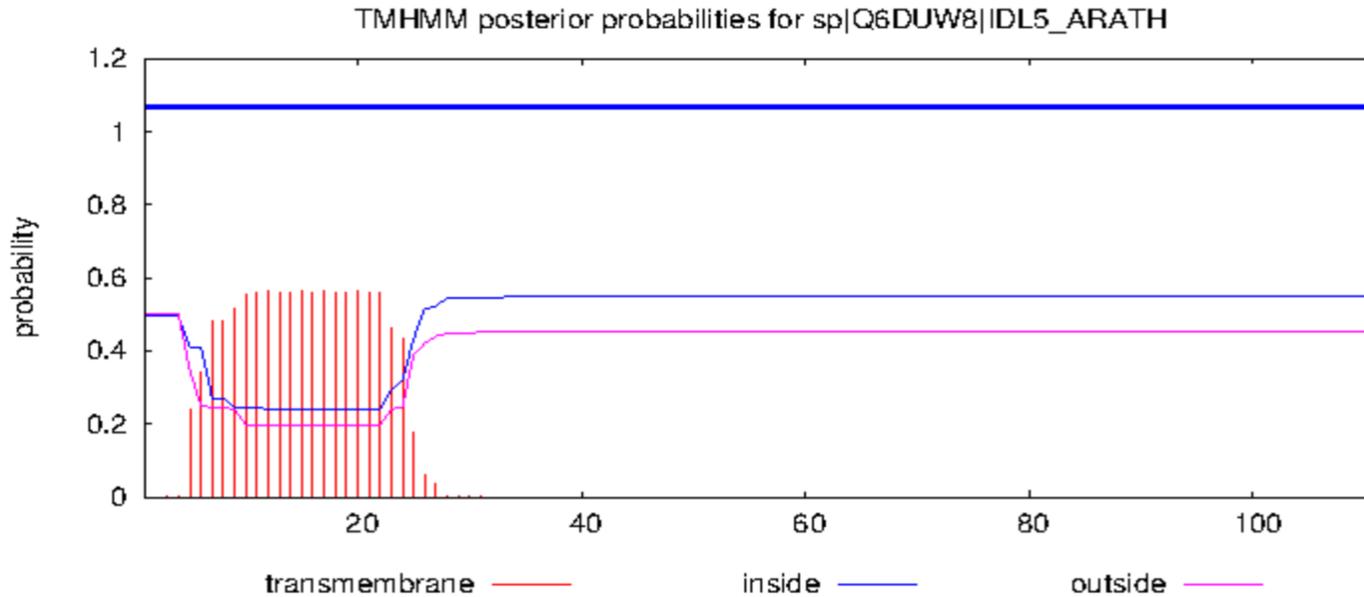
用Exrasy在线网站 (<https://web.expasy.org/protscale/>) 对其进行疏水性分析



纵坐标代表亲水性和疏水性系数，横坐标代表氨基酸序列；该小分子多肽属于亲水蛋白。

(二) 结构预测分析

通过TMHMM网站预测蛋白质跨膜结构域



纵坐标代表可能性，横坐标是蛋白质序列的位点，
Transmembrane表示跨膜结构
Inside表示膜内结构
Outside表示膜外结构

从图中跨膜区域主要集中在前30个位点，
序列后半段在膜内的可能性较高

(二) 结构预测分析

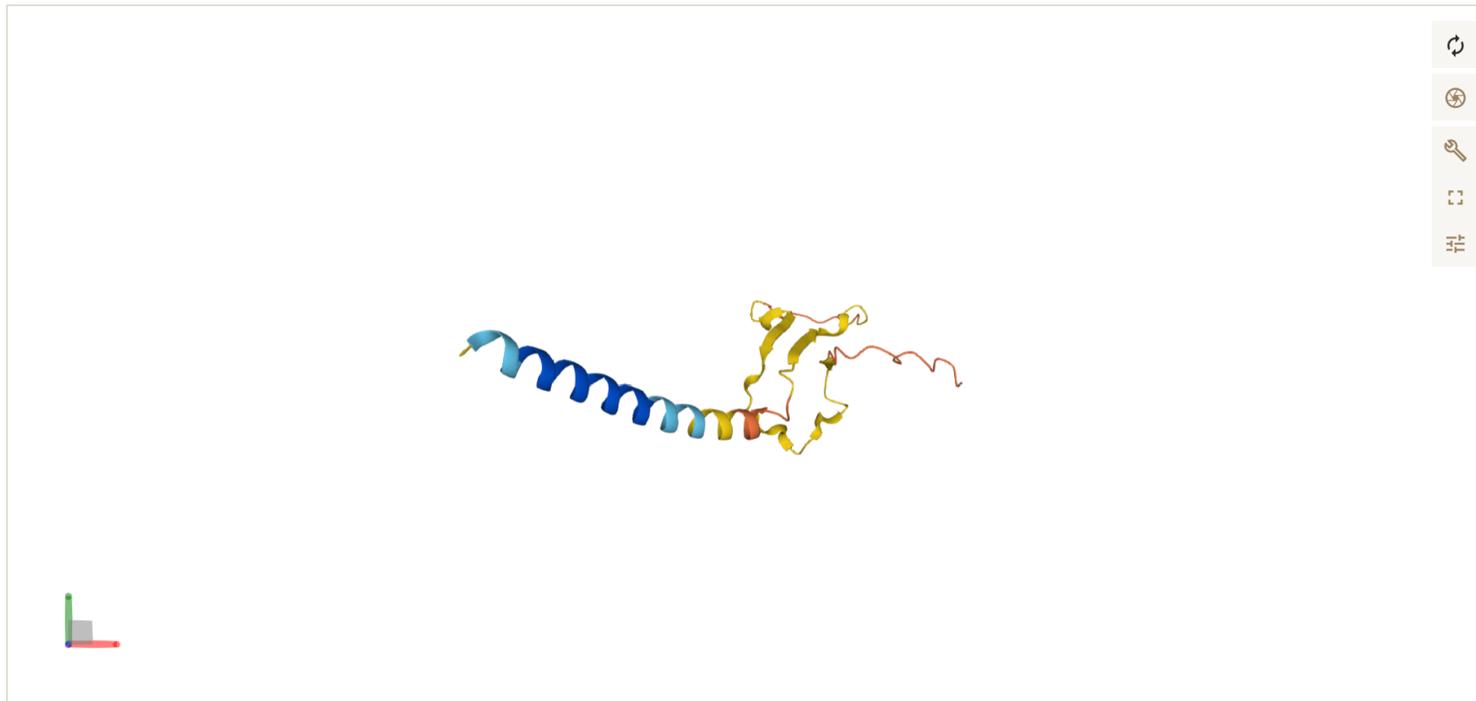
uniprot中的结构预测

Structureⁱ

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 with low pLDDT may be unstructured in isolation.



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS	
-- Select --	AF-Q6DUW8-F1	-- Select --			1-111	AlphaFold	



感谢老师和同学们的批评指正

THANK YOU FOR WATCHING