



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

新型CRISPR相关蛋白的初步鉴定

答辩人：钟玉楠

小组成员：颜瑞英 彭丹 曾欣



目录

- 1 背景介绍
- 2 初步验证



Part 1

背景介绍

一、基因编辑技术定义与发展历史



定义

基因编辑是一种精准化的基因工程新技术，指在生物体基因组水平上，借助经基因工程改造的**核酸酶**（“分子剪刀”）、**DNA识别域**（“基因组 GPS”）及细胞内天然的**DNA断点修复机制**（“基因组针线”），对**DNA核苷酸序列进行定点修饰**的过程。

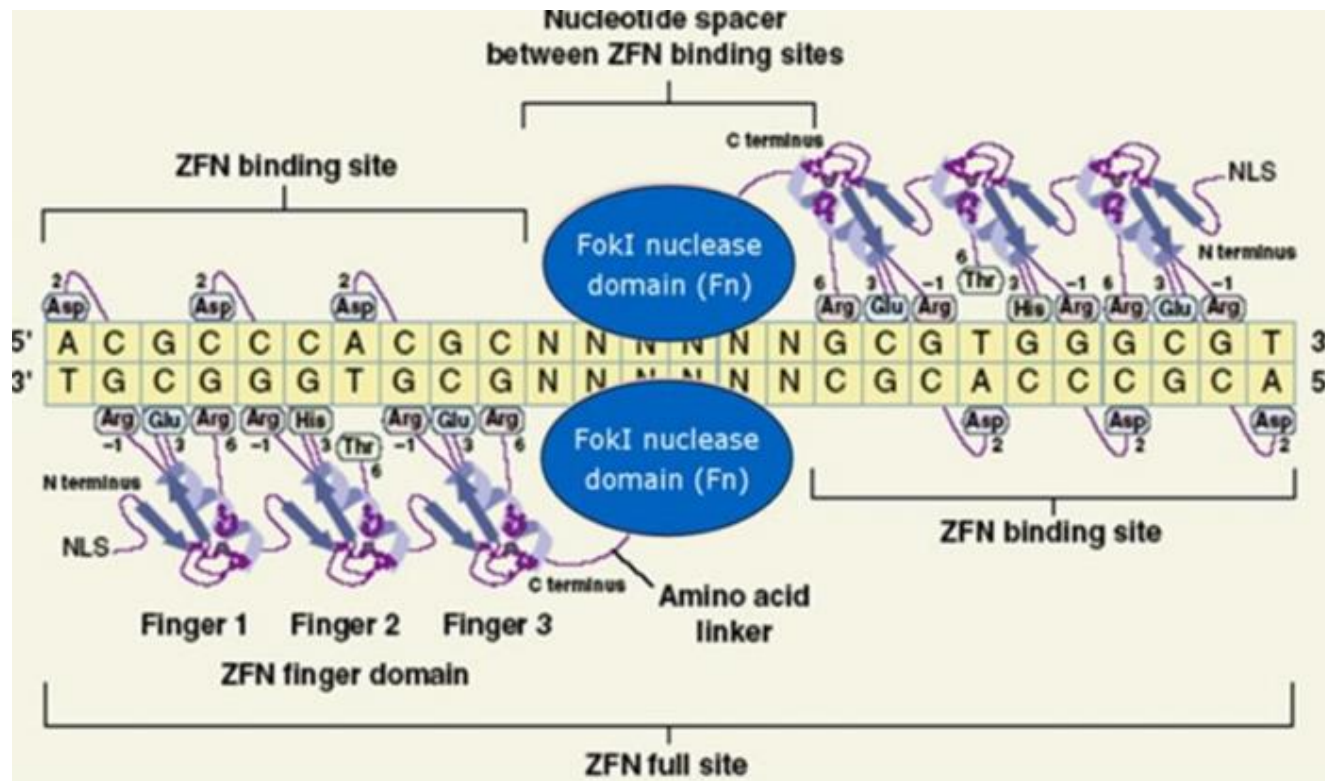
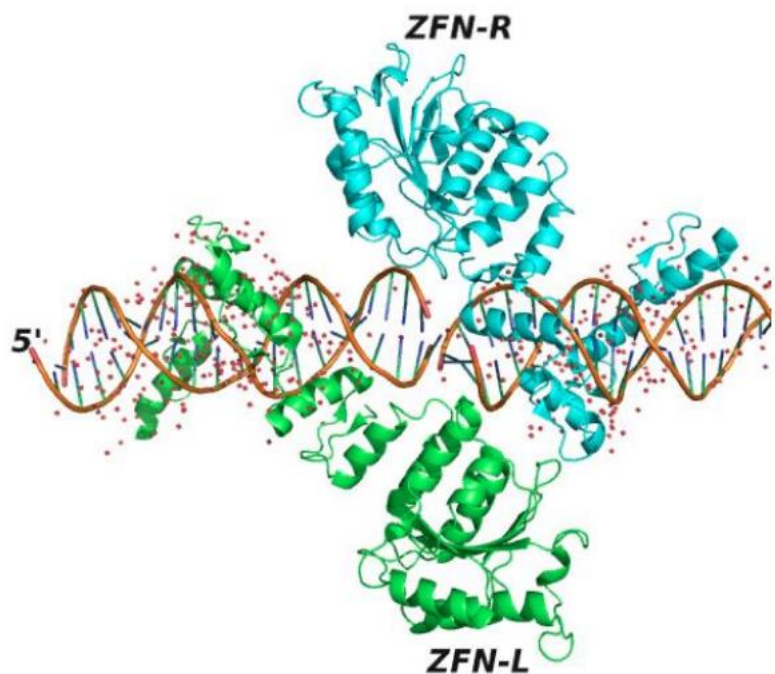
核心

通过特异性识别基因组靶位点并产生**双链断裂**（DSB），或在不产生双链断裂的情况下，利用**碱基修饰酶**等工具，引导细胞通过**非同源末端连接**（NHEJ）、**同源重组修复**（HDR）等机制，实现对特定DNA片段的**删除、插入、替换或单碱基转换**，模拟基因自然突变过程以修改原有基因组，最终精准**调控目标基因功能**，是研究基因生物功能、改良生物性状及治疗遗传疾病的重要工具。

第一代: ZFN



锌指核酸酶(Zinc-finger nucleases,ZFN)是人工改造的限制性核酸内切酶,将锌指蛋白与核酸内切酶FokI融合形成的**核酸内切酶**,利用不同的锌指结构**识别特异DNA序列**,利用核酸酶**切断靶DNA**。

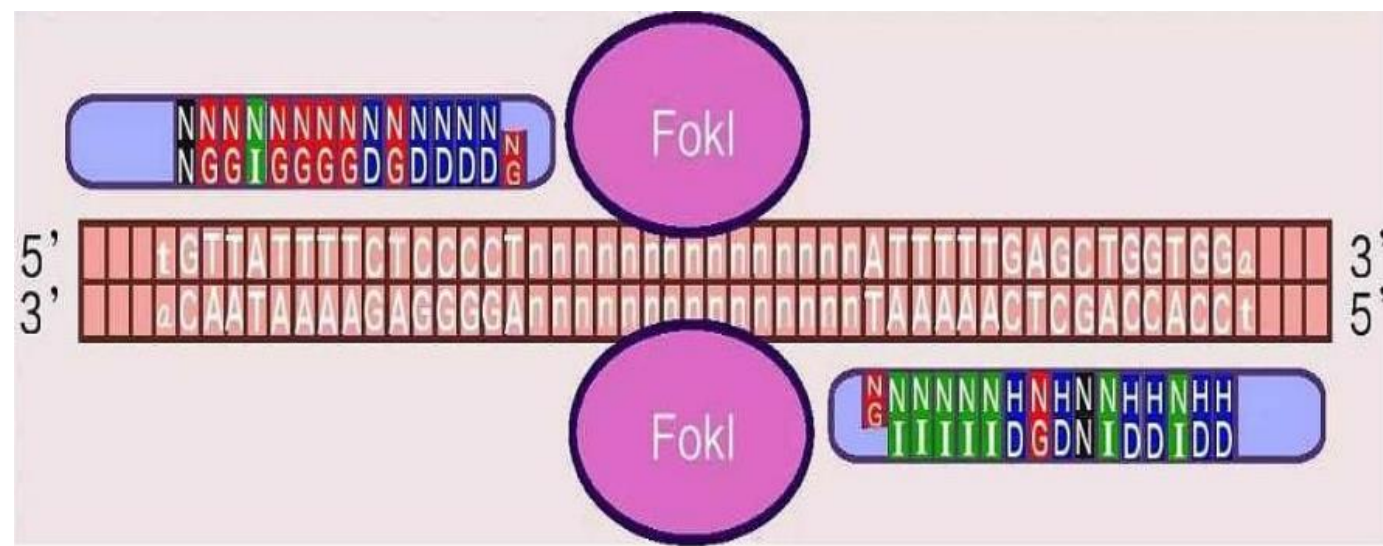
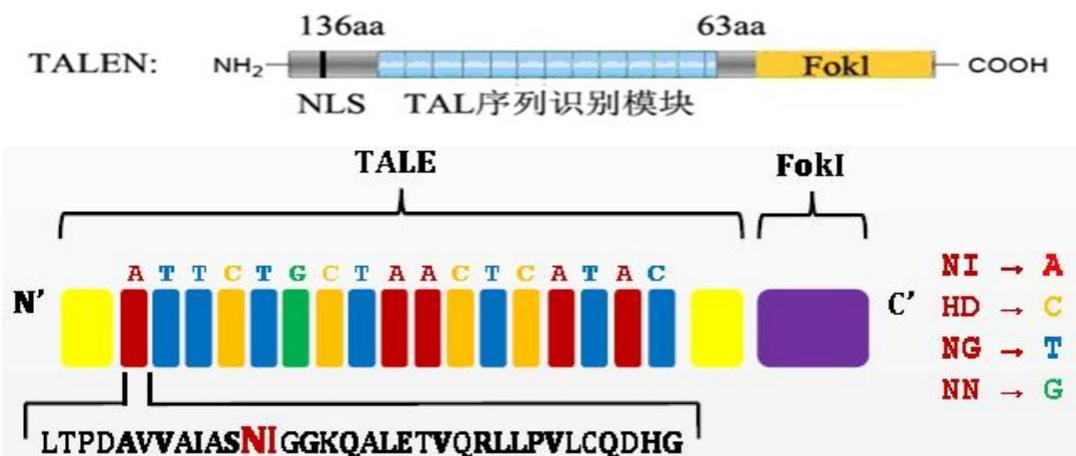


缺陷: DNA 识别序列长度有限; 易形成异源二聚体引发脱靶, 进而导致 DNA 错配、细胞毒性甚至细胞凋亡; 操作精准度不足, 可能诱发基因突变、癌症等风险; 体内使用易引发免疫反应且无法预测; 仅能用于体外细胞操作。

第二代: TALEN



类转录激活因子效应核酸酶(transcription activator-like effector nuclease, TALEN), TALEN可以对复杂的基因组进行**精细修饰**, 其**构建简单**, **特异性更高**, 每一根TALE对应一个DNA碱基, 而DNA一共只有4种碱基, 理论上只需要4根不同的TALE就可以把玩千变万化的万花筒游戏。



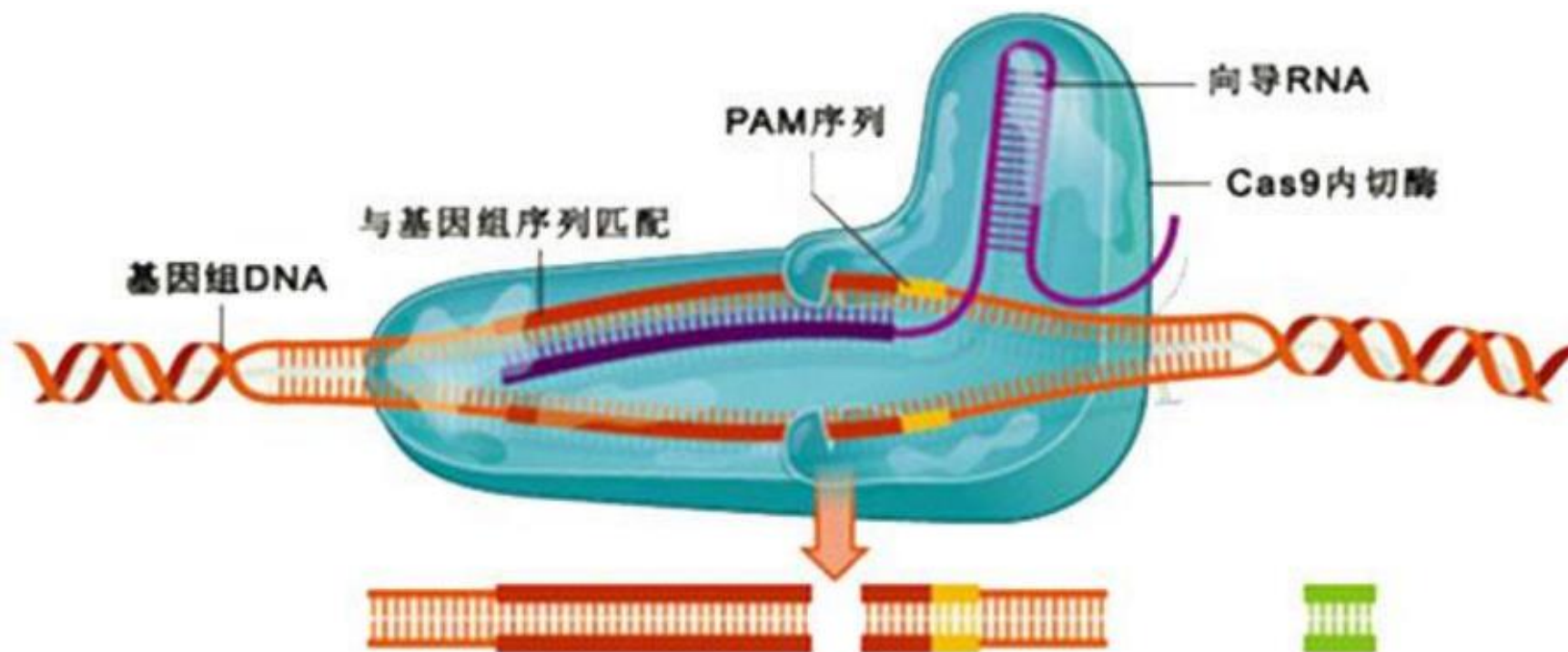
缺陷: 构建过程需组装多个重复模块, 操作较繁琐; TALE 蛋白分子量较大, 细胞内递送难度较高; 靶点序列要求长, 选择灵活性低; 批量构建成本高, 难高通量应用。

第三代：CRISPR/Cas



中国农业科学院
CHINESE ACADEMY OF AGRICULTURAL SCIENCES

成簇的规律间隔的短回文重复序列/Cas蛋白CRISPR-Cas9(clustered regular lyinterspaced short pal indromic repeats (CRISPR)/CRISPR-associated (Cas)system),发现于细菌和古细菌,基于细菌获得性免疫系统改造而成,CRISPR序列特异性**识别外源核酸序列**; Cas蛋白的非特异性核酸内切酶功能将**外源核酸切断**。



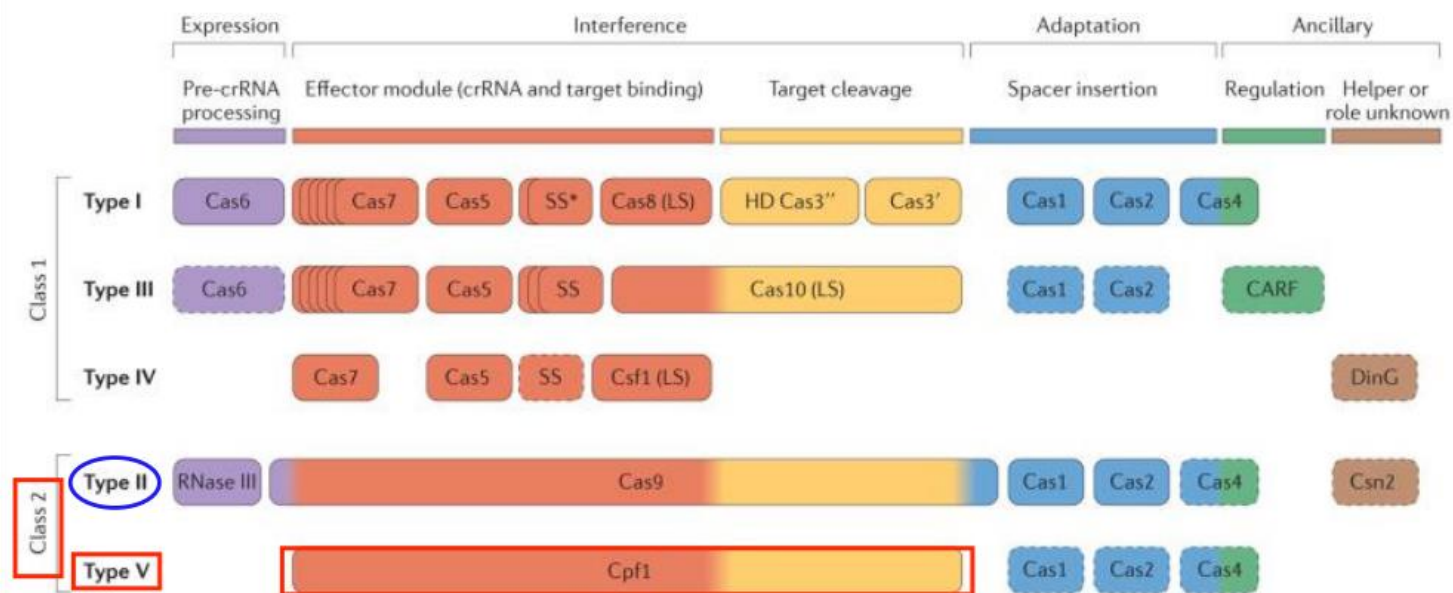
缺陷： PAM 序列限制严（仅识别 5'-NGG-3'），靶点覆盖范围窄；需 tracrRNA 辅助，向导 RNA 结构更复杂；切割产生平末端，外源基因定向插入效率低；无法自主加工 pre-crRNA，多基因编辑操作更繁琐。

二、Cas12a 的天然起源与分类定位



Cas12a (原称 Cpf1) 源于细菌 (如新弗朗西斯菌) 和古细菌的免疫防御系统, 通过识别噬菌体 DNA 的 PAM 序列并切割, 实现对二次入侵的抵御。

依据效应模块的组成方式分为2类: Class 1、Class 2



Cas12a (原 Cpf1) 属于 Class 2, 与 Cas9 并列, 是该类中首个被广泛开发的替代工具。核心优势是无需 tracrRNA、可自主加工 pre-crRNA, 为多基因编辑和高效递送提供基础。

流程图





Part **2**

初步验证

[Edit Search](#) Save Search Search Summary ▾

[How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title	Protein Sequence
RID	NU27V1MY014 Search expires on 01-07 20:07 pm Download All ▾
Program	BLASTP Citation ▾
Database	ClusteredNR See details ▾
Query ID	lcl Query_3932045
Description	unnamed protein product
Molecule type	amino acid
Query Length	1334
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results

Organism only top 20 will appear **NEW**

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

Clusters Graphic Summary Alignments Taxonomy

Clusters producing significant alignments

Download ▾ Select columns ▾ Show 100 ▾ [?](#)

select all 100 clusters selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

	Cluster Composition	Cluster Ancestor	Cluster Representative Sequence	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	g-proteobacteria	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Succinivibrio dextrino...	2685	2685	100%	0.0	100.00%	1334	WP_031492824.1
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	g-proteobacteria	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Succinivibrio sp.]	728	728	50%	0.0	57.86%	682	MGN1280677.1
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	g-proteobacteria	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Succinivibrionaceae ...	1111	1111	99%	0.0	46.87%	1301	MGN1392737.1
<input checked="" type="checkbox"/>	3 member(s), 1 organism(s)	g-proteobacteria	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Succinivibrio sp.]	1116	1116	100%	0.0	46.11%	1301	MDD6068465.1
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	g-proteobacteria	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Succinatimonas sp.]	1068	1068	100%	0.0	45.36%	1302	MCI6345495.1
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	g-proteobacteria	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Ruminobacter sp.]	961	961	100%	0.0	43.36%	1328	MBR1924181.1
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	firmicutes	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Phascolarctobacteriu...	896	896	100%	0.0	41.26%	1297	MBQ9764161.1
<input checked="" type="checkbox"/>	1 member(s), 1 organism(s)	firmicutes	type.V.CRISPR-associated_protein_Cas12a/Cpf1 [Anaerovoracaceae b...	853	853	100%	0.0	39.62%	1301	MEF9921705.1

根据NCBI blast结果可知，我们的未知序列匹配上了**Cas12a序列**，比对上的区域中序列一致性最高为**57.86%**

III InterPro

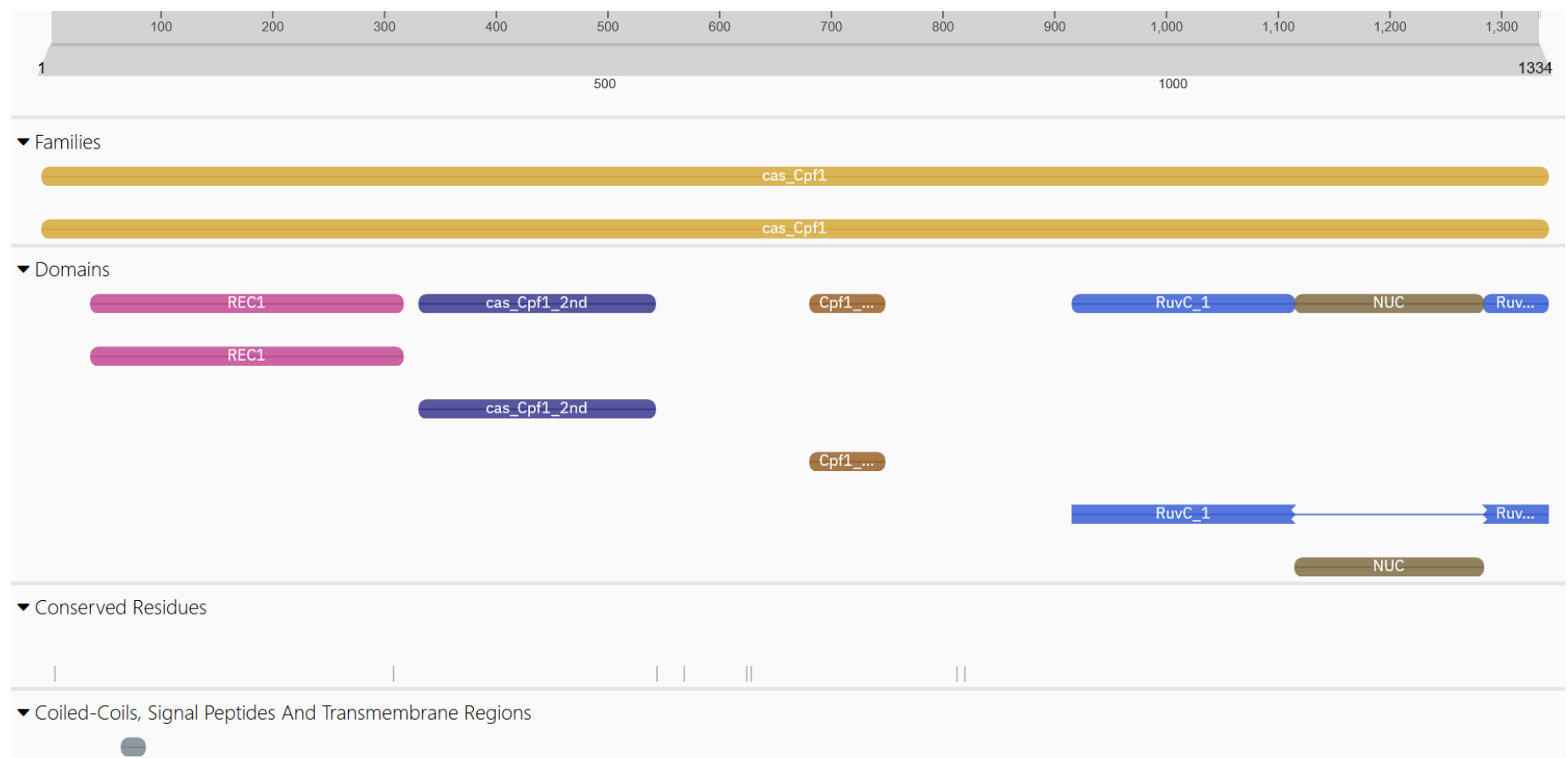


利用InterPro做解析序列的结构域组成:

发现未知序列有Cas12a典型的REC1、REC2、PI和NUC结构域, 并且含有RUVVC结构域

Entry matches to this protein¹

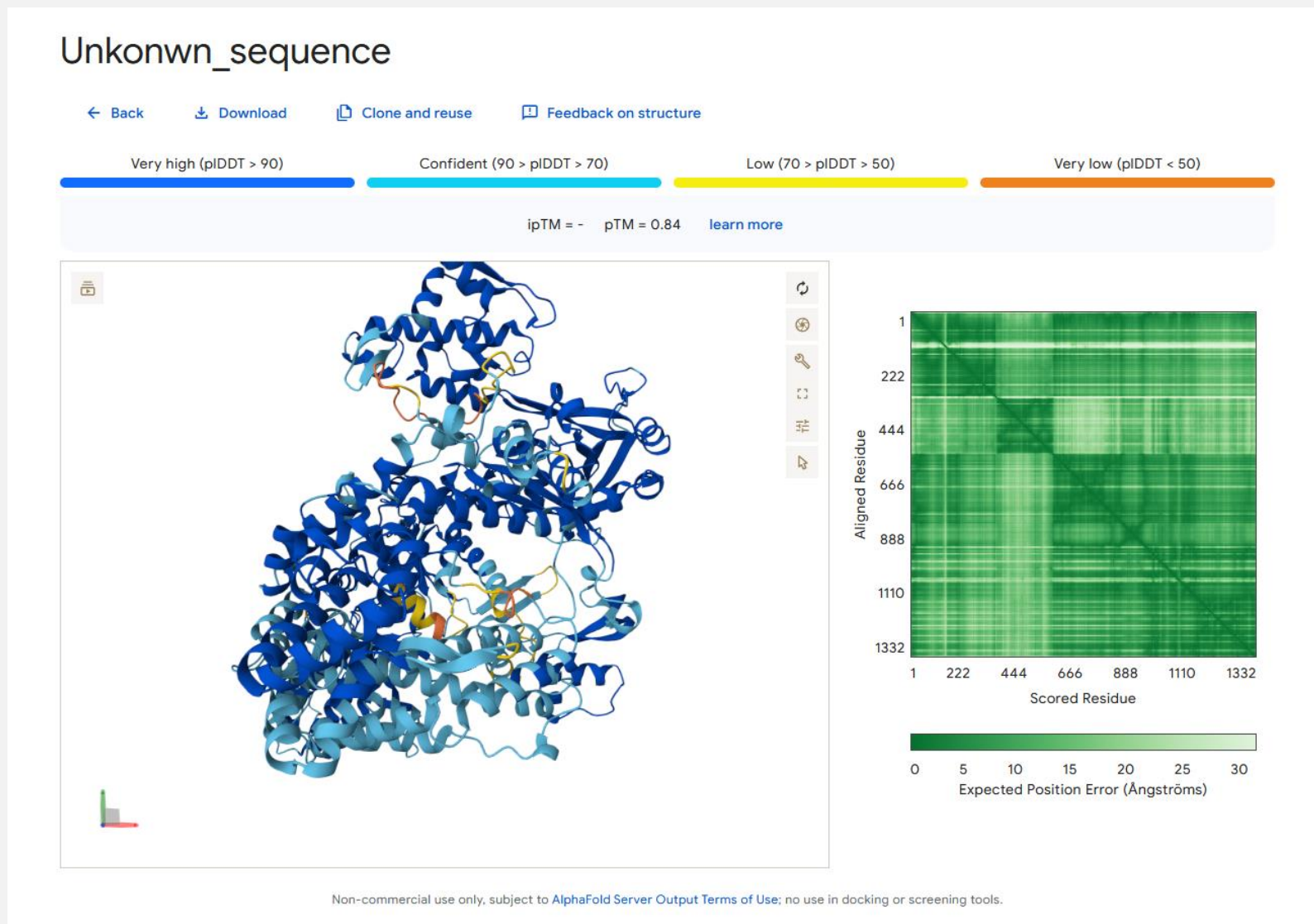
Options Download Feature Display Mode ?
 Summary Full



V AlphaFold



中国农业科学院
CHINESE ACADEMY OF AGRICULTURAL SCIENCES



AlphaFold预测未知序列结构，ipTM = -pTM = 0.84，说明预测的蛋白结构质量较高

VI Pymol

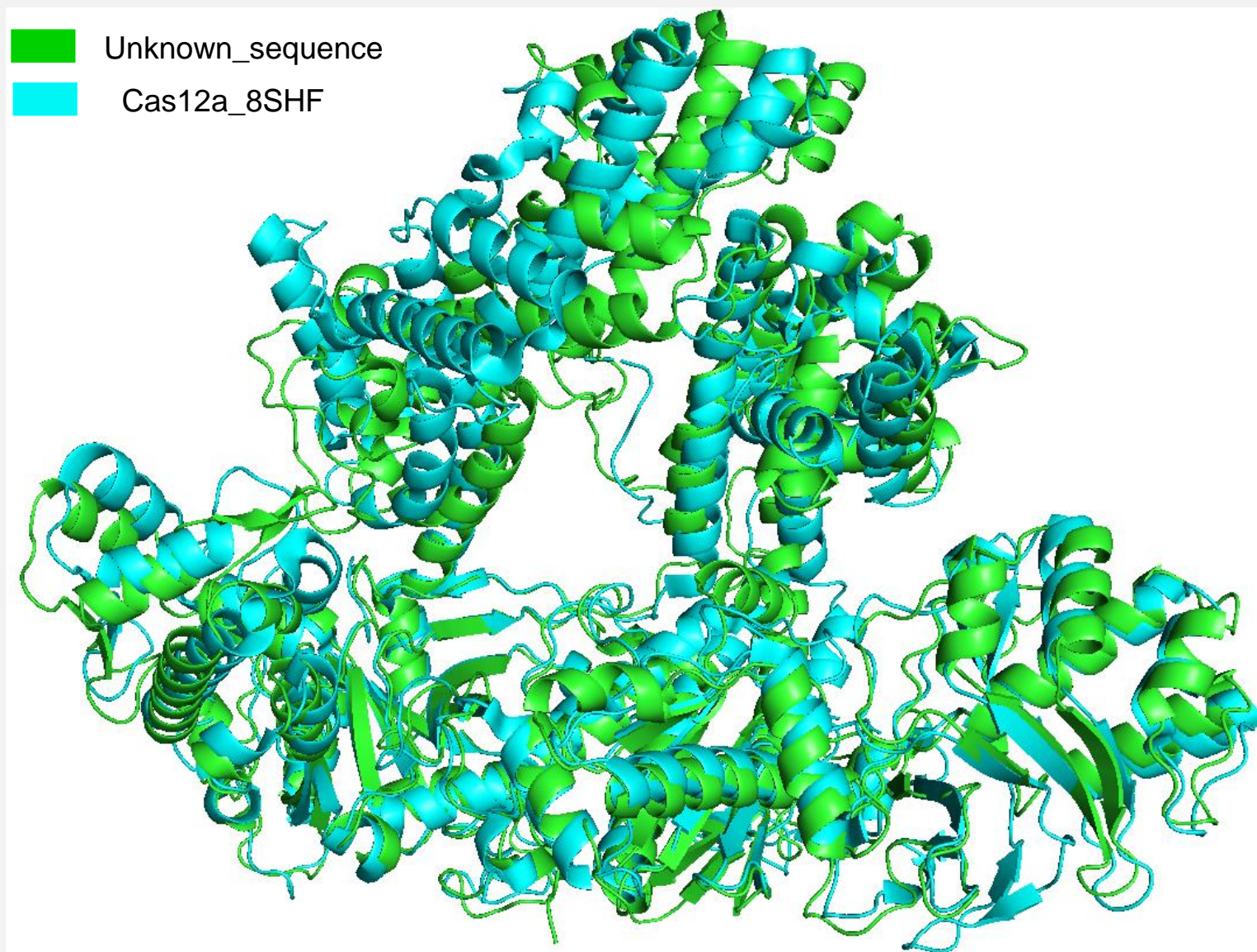


中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

将预测的未知序列与已发表的Cas12a蛋白结构8SFH进行结构叠加
super(RMSD)结果为2.319，说明两个蛋白的结构相似性很高

```
PyMOL>super unknown_sequence, 8SFH, object="2"  
MatchAlign: aligning residues (1334 vs 1398)...  
MatchAlign: score 4097.532  
ExecutiveAlign: 7881 atoms aligned.  
ExecutiveRMS: 475 atoms rejected during cycle 1 (RMSD=4.99).  
ExecutiveRMS: 432 atoms rejected during cycle 2 (RMSD=3.93).  
ExecutiveRMS: 396 atoms rejected during cycle 3 (RMSD=3.27).  
ExecutiveRMS: 311 atoms rejected during cycle 4 (RMSD=2.80).  
ExecutiveRMS: 243 atoms rejected during cycle 5 (RMSD=2.51).  
Executive: RMSD = 2.319 (6024 to 6024 atoms)
```



VII EMBOSS



	A	B	C
1	Query_ID	Target_ID	Identity (%)
2	Unkonwn_sequence	WP_031492824.1	100
3	Unkonwn_sequence	MGN1392737.1	46.6
4	Unkonwn_sequence	MDD6068465.1	45.9
5	Unkonwn_sequence	MCI6345495.1	45.1
6	Unkonwn_sequence	MBR1924181.1	43.3
7	Unkonwn_sequence	MBQ9764161.1	41.4
8	Unkonwn_sequence	WP_407399469.1	39.7
9	Unkonwn_sequence	MEG1642078.1	39.5
10	Unkonwn_sequence	MEF9921705.1	39.4
11	Unkonwn_sequence	WP_295861397.1	39.3
12	Unkonwn_sequence	WP_318697387.1	39.3
13	Unkonwn_sequence	NLM08782.1	39.2
14	Unkonwn_sequence	MD09155116.1	39.1
15	Unkonwn_sequence	MBD5439500.1	38.8
16	Unkonwn_sequence	MCQ2584096.1	38.5
17	Unkonwn_sequence	WP_318675263.1	38.4
18	Unkonwn_sequence	MCH5150260.1	38.4
19	Unkonwn_sequence	WP_295359647.1	38.4
20	Unkonwn_sequence	MBD5411347.1	38.3
21	Unkonwn_sequence	MEE1269210.1	38.2
22	Unkonwn_sequence	MBR1591151.1	38.1
23	Unkonwn_sequence	WP_370821459.1	38.1
24	Unkonwn_sequence	MCI7578025.1	38
25	Unkonwn_sequence	WP_390603654.1	37.9
26	Unkonwn_sequence	Cas12a_8SFH	37.8
27	Unkonwn_sequence	MBP5451644.1	37.7
28	Unkonwn_sequence	WP_184651713.1	37.6
29	Unkonwn_sequence	MBR1537826.1	37.6
30	Unkonwn_sequence	MBO4859624.1	37.6
31	Unkonwn_sequence	WP_290958800.1	37.6
32	Unkonwn_sequence	MCQ2588525.1	37.5

利用EMBOSS中的`needle`工具，将这个未知序列与我们收集的序列做**全局比对**，将比对结果整理成EXCEL表格，由右图可以看到，除了序列本身外，其他比对的结果都在**50%以下**，说明这个未知序列可能是一个**新的Cas12a序列**

下一步实验验证



- 1、大肠杆菌验证基因编辑活性 (PAM)
- 2、动物细胞验证基因编辑活性 (主要)
- 3、植物验证基因编辑活性

成员介绍



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



C



D



B



A

成员

工作分配

A-颜瑞英

背景介绍、PPT美化

B-彭丹

工具查找、第一部分

C-钟玉楠

期末汇报、第二部分

D-曾欣

资料查找、PPT排版



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

恳请大家批评指正!

