

# 基于结构的 CRISPR 蛋白 xCas9 的优化设计

薛冬梅, 朱海霞, 杜文豪, 汤洪海, 黄强

复旦大学 生命科学学院, 上海 200438

薛冬梅, 朱海霞, 杜文豪, 等. 基于结构的 CRISPR 蛋白 xCas9 的优化设计. 生物工程学报, 2021, 37(4): 1385-1395.

Xue DM, Zhu HX, Du WH, et al. Structure-based optimization and design of CRISPR protein xCas9. Chin J Biotech, 2021, 37(4): 1385-1395.

**摘要:** 酿脓链球菌 (*Streptococcus pyogenes* Cas9, SpCas9) 已成为强大的基因组编辑工具, 但其可识别的前间隔序列临近基序 (Protospacer adjacent motifs, PAMs) 范围有限, 且存在脱靶效应。为解决这些问题, 文中提出一种对 SpCas9 的定向进化突变体 xCas9 进行优化的理性方法。首先, 使用 Rosetta 程序进行能量最小化以优化 Cas9 的三维结构, 获得其能量最低的构象; 然后, 对其定向进化所得氨基酸位点进行组合突变设计; 最后, 通过自由能排序从设计突变体中筛选出用于实验验证的最优突变体。经 DNA 剪切实验验证, 成功地获得了一个多 PAM 识别和低脱靶的新突变体 yCas9 (262A/324R/409N/480K/543D/694L/1219T)。该突变体可识别 NG、GAA 和 GAT 序列, 且其由错配 sgRNA 引导的脱靶 DNA 剪切活性低, 为生物医学领域提供了一个有潜在应用价值的基因编辑工具。同时, 文中还对 SpCas9、xCas9 和 yCas9 进行了分子动力学模拟, 揭示了其 PAM 识别和脱靶效应的机理, 可为进一步的 CRISPR/Cas9 蛋白优化改造提供理论指导。

**关键词:** 基因编辑, CRISPR/Cas9, 蛋白计算设计, 分子动力学模拟

## Structure-based optimization and design of CRISPR protein xCas9

Dongmei Xue, Haixia Zhu, Wenhao Du, Honghai Tang, and Qiang Huang

School of Life Sciences, Fudan University, Shanghai 200438, China

**Abstract:** *Streptococcus pyogenes* Cas9 (SpCas9) has become a powerful genome editing tool, but has a limited range of recognizable protospacer adjacent motifs (PAMs) and shows off-target effects. To address these issues, we present a rational approach to optimize the xCas9 mutant derived from SpCas9 by directed evolution. Firstly, energy minimization with the Rosetta program was applied to optimize the three-dimensional structure of Cas9 to obtain the lowest energy conformation. Subsequently, combinatorial mutations were designed based on the mutations sites of xCas9 acquired during the directed

**Received:** June 30, 2020; **Accepted:** September 27, 2020

**Supported by:** National Major Scientific and Technological Special Project for 'Significant New Drugs Development' (No. 2018ZX09J18112), National Natural Science Foundation of China (Nos. 31671386, 31971377).

**Corresponding author:** Qiang Huang. Tel: +86-21-31246589; E-mail: huangqiang@fudan.edu.cn

国家重大科技专项“重大新药开发”课题 (No. 2018ZX09J18112), 国家自然科学基金 (Nos. 31671386, 31971377) 资助。

网络出版时间: 2020-10-13

网络出版地址: <https://kns.cnki.net/kcms/detail/11.1998.Q.20201013.0948.005.html>

evolution. Finally, optimal mutants were selected from the designed mutants by free energy ranking and subjected to experimental verification. A new mutant yCas9 (262A/324R/409N/480K/543D/694L/1219T) with multiple PAM recognition ability and low off-target effects was obtained and verified by DNA cleavage experiments. This mutant recognizes the NG, GAA and GAT PAMs and shows low off-target DNA cleavage activity guided by mismatched sgRNA, thus provides a gene editing tool with potential applications in biomedical field. Furthermore, we performed molecular dynamics simulations on the structures of SpCas9, xCas9 and yCas9 to reveal the mechanisms of their PAM recognition and off-target effects. These may provide theoretical guidance for further optimization and modification of CRISPR/Cas9 proteins.

**Keywords:** gene editing, CRISPR/Cas9, computational protein design, molecular dynamics simulation

CRISPR 是细菌和古菌中抵抗外源核酸入侵的免疫系统。近年来,该系统已被人们开发成一种特异性剪切 DNA 的基因编辑工具<sup>[1]</sup>。目前最常用的是 II 型 CRISPR 系统中的酿脓链球菌 Cas9 (*Streptococcus pyogenes* Cas9, SpCas9)<sup>[2]</sup>。不过, SpCas9 存在前间隔序列临近基序 (Protospacer adjacent motifs, PAMs) 识别范围有限和易脱靶的问题,这限制了其广泛应用<sup>[3-5]</sup>。因此,只有拓展 PAM 识别范围和降低脱靶效应,才能使 SpCas9 技术在基因编辑领域发挥更大的作用。

为了达到上述目标,人们开发了多种 SpCas9 的优化方法。例如,利用细菌选择系统来筛选设计 SpCas9,通过该方法获得了 PAM 识别范围扩大的 VQR、EQR 和 VRER 突变体<sup>[6]</sup>;还有,基于 SpCas9 结构的改造方法,也获得了高保真突变体 eSpCas9、SpCas9-HF、HypaCas9 和 evoCas9<sup>[7-8]</sup>。此外,David Liu 研究组通过 PACE 定向进化法获得了一个新型 SpCas9 突变体 (xCas9),与野生型相比,该突变体具有更高的特异性,且可以识别包括 NGG、NGA、NGT、NGC、GAA 和 GAT 在内的广泛的 PAM 序列<sup>[9]</sup>。但是,xCas9 在提高特异性和 PAM 兼容性的同时,其 DNA 剪切活性有较大损失,对不同 PAM 的识别能力也存在差异<sup>[10-11]</sup>。因此,该蛋白还有进一步优化的空间。

然而,上述几种优化 SpCas9 的方法有一定的盲目性,耗时耗力。在本研究中,我们提出了一种新的优化 SpCas9 的方法。该方法从能量的角度出发,用 Rosetta 程序对定向进化突变体 xCas9 进行优化。Rosetta 是一个精确的生物大分子设计

程序,可通过突变 DNA 周围的关键氨基酸来优化蛋白-DNA 的相互作用界面,已在蛋白-DNA 体系中得到了成功应用<sup>[12]</sup>。鉴于 Cas9 剪切 DNA 属于蛋白-DNA 相互作用的一种,我们设想,利用 Rosetta 来突变对 Cas9 蛋白特性有重要影响的氨基酸,以优化 Cas9-DNA 的相互作用界面,获得自由能较低的突变体,这些突变体极有可能具有高 PAM 兼容性或低脱靶率的属性。

本研究以 xCas9 为对象,选取其分布在 sgRNA-DNA 异源双链体两侧的 7 个定向进化位点<sup>[9]</sup>作为优化设计的目标位点(图 1),并对这些位点进行 20 种氨基酸的组合设计,最终获得了一个突变体 yCas9。该突变体具有高 PAM 兼容性,可识别 NGG、NGA、NGT、NGC、GAA 和 GAT 序列,且对 GAT PAM 的识别能力比 xCas9 高。同时,与野生型 SpCas9 相比,yCas9 的脱靶效应也大幅度降低。同时,我们还通过分子动力学模拟分析了 SpCas9 的 PAM 识别和脱靶效应的分子机理,确定了 1 219 位氨基酸与 Cas9 的 PAM 识别之间的关系,以及 REC 域构象重排与 Cas9 脱靶效应之间的关系。

## 1 材料与方法

### 1.1 实验材料

本研究使用的 pET21-6His-TEV-SpCas9、pET21-6His-TEV-xCas9、psk-sgRNA 和 psk-DNA 质粒由本课题组保存。转化用的大肠杆菌 *Escherichia coli* DH5 $\alpha$  (CB101-02) 和 Rosetta (DE3) (CB108-02)、质粒提取试剂盒 (DP103-02)、点突

变试剂盒 (KM101) 等购自 TIANGEN 生化科技 (北京) 有限公司; 胶回收试剂盒 (AP-GX-50) 购自 Axygen 公司; 一步定向克隆试剂盒 (NR005-01A) 购自近岸蛋白质科技公司; PrimeSTAR GXL DNA Polymerase (R050Q) 购自 TaKaRa 公司; RNA 转

录试剂盒 (AM1354) 和 RNA 纯化试剂盒 (AM1908) 购自 Thermo Fisher 公司。本研究所用引物均采用软件 Primer 5 设计, 满足各试剂盒的引物设计原则, 并委托上海杰李生物技术有限公司合成 (表 1)。

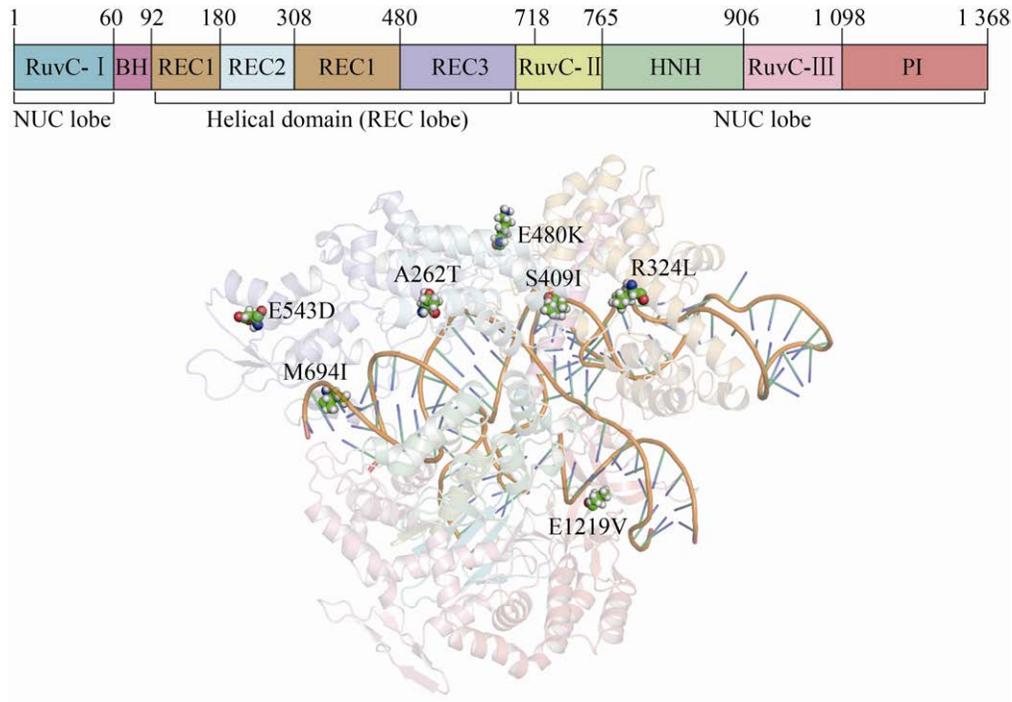


图 1 xCas9 的 7 个定向进化位点<sup>[9]</sup>的三维分布图

Fig. 1 3D-distribution of seven directed-evolution sites of xCas9<sup>[9]</sup>.

表 1 本研究所用引物

Table 1 Primers used in this study

Primer name	Sequence (5'-3')	Usage
NKDL-LF	CTGCAGTTGATCCATGATGACTCTCTCAC	Mutations of 409, 480, 543 and 694
NKDL-LR	ATTCCATTGTCGAAAGTGCGCTGTTTCG	
NKDL-SF	GCAAACAGCGCACTTTCGACAATGGAAATATCCCCACCAGATCACCTGGGC	
NKDL-SR	GTGAGAGAGTCATCATGGATCAACTGCAGGAAGTCCGGTTGGCAAATCCATCG	
E1219T-F	CTCGCTAGTGCGGGCACCCTGCAGAAAGGTAACG	Mutation of 1 219
E1219T-R	CGTTACCTTTCTGCAGGGTGCCCGCACTAGCGAG	
DNA-F	TAGTCCTGTCGGGTTTCG	Amplification of target DNA (920 bp)
DNA-R	TTCCATTGCCATTCAGG	
sgRNA-F	GGAATTCGATATCAAGCTTA	Amplification of sgRNA (97 nt)
sgRNA-R	AGCACCGACTCGGTGCCACTT	

## 1.2 xCas9 的组合设计

生物分子结构预测和设计的 Rosetta 程序包含 100 多个模块,具有蛋白质结构从头预测、蛋白质结构设计或对接以及蛋白质重设计等功能<sup>[13-14]</sup>。为了实现这些功能, Rosetta 通过其独特的自由能打分函数计算蛋白质能量,包括不同残基间的引力和斥力、蛋白骨架与侧链以及侧链与侧链之间的氢键相互作用、范德华作用等。当使用该打分函数比较同一分子的不同构象时,得分低的构象更稳定,自由能更低,更接近于天然结构。

本文使用了 Rosetta 程序的两个模块: Relax 和 Fixed backbone design (Fixbb)。Relax 模块通过多次迭代使初始结构的氨基酸侧链重排,并且在每轮迭代中不断地调整分子间作用力,以搜索该结构的最优构象。最终经过 Relax 模块处理后的结构,一般都处于能量最低的构象。Fixbb 模块在固定蛋白骨架的同时,允许氨基酸的侧链类型自由设计,以设计与原蛋白骨架兼容的氨基酸侧链,最后通过自由能排序获得最优突变体。

本文设计 Cas9 突变体的主要流程是:首先从蛋白质数据库 (<http://www.rcsb.org/>) 下载 SpCas9 结构 (PDB ID: 4UN3)。其次,用 Relax 模块将该结构进行 25 轮迭代,并根据 Rosetta 的自由能打分函数选出能量最低的结构作为初始结构。然后,使用 Fixbb 模块固定 Cas9 蛋白的主链原子,并对 xCas9 的 7 个定向进化位点 (262、324、409、480、543、694、1 219) 进行 20 种氨基酸组合设计。最后,进行 10 000 次独立设计,并对输出结果进行去重处理和按自由能大小排列筛选。

## 1.3 蛋白突变体质粒的构造

本文分别以质粒 pET21-6His-TEV-SpCas9 和 pET21-6His-TEV-xCas9 为模板,通过一步定向克隆和点突变的方式构建突变体质粒。

## 1.4 蛋白的表达与纯化

将 Cas9 及其突变体质粒 (N 端带有 His-tag) 转入 *E. coli* Rosetta(DE3) 细胞中表达,当菌液

的  $OD_{600}$  达到 0.9 时,加入终浓度为 0.5  $\mu\text{mol/L}$  的异丙基- $\beta$ -D-硫代半乳糖苷 (Isopropyl- $\beta$ -D-thiogalactopyranoside, IPTG), 16  $^{\circ}\text{C}$  过夜培养。离心收集菌体,并用裂解液 (20 mmol/L HEPES, 500 mmol/L KCl, 0.1 mmol/L PMSF, pH 7.5) 重悬菌体,超声破碎 20 min, 12 000 r/min 离心 1 h。上清液与 Ni-NTA 孵育 90 min 后,用缓冲液 B (20 mmol/L HEPES, 400 mmol/L KCl, 500 mmol/L 咪唑, 1%蔗糖, pH 7.5) 梯度洗脱 Cas9 蛋白。经过 SDS-PAGE 鉴定后,收集目的蛋白并将其浓缩至 250  $\mu\text{L}$ ,保存于蛋白储存液 (20 mmol/L HEPES, 500 mmol/L KCl, 1 mmol/L DTT, pH 7.5) 中。

## 1.5 底物 DNA 的获取

以 psk-DNA 质粒为模板,使用引物 DNA-F/R 对包含 PAM 序列和 20 bp 靶序列的 DNA 进行 PCR 扩增,然后切胶回收获得 920 bp 的底物 DNA,并将其保存于无核酸酶的水中。DNA 包含不同的 PAM 序列,分别为 5'-TGG-3'、5'-TGA-3'、5'-TGC-3'、5'-TGT-3'、5'-GAA-3'和 5'-GAT-3'。

## 1.6 sgRNA 的获取

首先以 psk-sgRNA 质粒为模板,使用引物 sgRNA-F/R 进行 PCR 扩增。其次,切胶回收获得 sgRNA 的转录模板,并使用 RNA 转录试剂盒在体外将模板转录成 sgRNA, 37  $^{\circ}\text{C}$  温育 6 h。然后,用 TURBO DNase 消化转录模板。最后,用 RNA 纯化试剂盒纯化产物,并将其保存于无核酸酶的水中。sgRNA 中与底物 DNA 的靶序列匹配的间隔序列为 5'-UACCGCUCCAGUCGUUCAUG-3'。用于脱靶检测的 sgRNA 构建过程:从 PAM 远端开始依次将 sgRNA 间隔序列的两个碱基突变为其互补碱基,共改 8 次,最终获得 9 个用于脱靶检测的 sgRNA,其序列如表 2 所示。

## 1.7 体外剪切活性检测

本文构建的体外剪切活性检测反应体系为 20  $\mu\text{L}$  体系,包含 200 nmol/L Cas9、200 nmol/L sgRNA、30 nmol/L DNA 和反应缓冲液 (20 mmol/L

表 2 sgRNA 序列

Table 2 Sequences of sgRNA

No.	Sequence (5'-3')	Site
0	UACCGCUCCAGUCGUUCAUG	
1	<u>A</u> UCCGCUCCAGUCGUUCAUG	01-02
2	U <u>U</u> GCGCUCCAGUCGUUCAUG	02-03
3	UAG <u>G</u> GCUCAGUCGUUCAUG	03-04
4	UAC <u>G</u> CCUCCAGUCGUUCAUG	04-05
5	UACCC <u>G</u> UCCAGUCGUUCAUG	05-06
6	UACCC <u>G</u> <u>A</u> CCAGUCGUUCAUG	06-07
7	UACCGC <u>A</u> GCAGUCGUUCAUG	07-08
8	UACCGC <u>G</u> AGUCGUUCAUG	08-09

Note: the underline part of the sequence is mutated bases.

HEPES, 150 mmol/L KCl, 1 mmol/L DTT, 10 mmol/L MgCl<sub>2</sub>, pH 7.5)。首先在反应缓冲液中加入 Cas9 和 sgRNA,混合均匀后将其放置在 37 °C 预孵育 10 min,然后将 DNA 加入混合物中,继续孵育,孵育时间随检测目的而定(检测 PAM 识别时,时间设为 2 h;检测脱靶效应时,时间设为 3 h)。反应结束后,使用琼脂糖凝胶电泳检测 Cas9 的底物剪切情况。

### 1.8 分子动力学模拟

本文使用 GROMACS 对 Cas9 蛋白进行分子动力学 (Molecular dynamics, MD) 模拟,以获得其在溶液环境中的稳定构象<sup>[15]</sup>。从蛋白质数据库中下载 SpCas9 结构 (PDB ID: 4UN3) 和 xCas9 结构 (PDB IDs: 6K4P、6AEG)。晶体结构中缺失的氨基酸残基用在线网站 SWISS-MODEL (<https://swissmodel.expasy.org/>) 同源建模补齐, yCas9 的 3D 结构用 Pymol (<https://pymol.org/2/>) 点突变获得。使用 Amber ff14SB\_Parmbsc1 力场和 TIP3P 水模型对 Cas9 系统进行建模<sup>[16]</sup>。在模拟系统中, Cas9 的全原子结构被放在一个离盒壁至少  $1.5 \times 10^{-9}$  m 的立方体盒中。在该体系中添加一定数量的 Na<sup>+</sup>和 Cl<sup>-</sup>以中和系统的净电荷,并使离子浓度达到 0.15 mol/L。模拟过程中先进行能量最小化,使体系达到预平衡状态;接着,依次进行 3 次 NVT 平衡,分别为 100 K 平衡 200 ps、200 K 平衡 200 ps 和 300 K 平衡 200 ps,最终使系统温度保持在 300 K;然后,进行 2 ns 的 NPT 平衡,使

体系达到  $1 \times 10^{-5}$  Pa 的恒定压力;最后在 300 K、 $1 \times 10^{-5}$  Pa 的条件下进行 200 ns 的 MD 模拟,每 40 ps 输出一帧,供后续分析使用。

## 2 结果与分析

### 2.1 突变体的获得

本研究使用 Rosetta 的 Fixbb 模块对 xCas9 的 7 个定向进化位点进行了组合设计,共输出了 10 000 个独立的组合设计。根据蛋白质折叠的热力学假说,一个蛋白质构象的自由能越低,则该构象在自然界中出现的概率越大,构象越稳定<sup>[17-19]</sup>。所以,我们对这 10 000 个设计进行去重处理和排序后,最终获得两个能量最低的突变体(表 3)。如表 3 所示,这两个突变体的 Rosetta 自由能打分函数的值分别为 -1 704.054 和 -1 703.748。因此,在 xCas9 定向进化位点的组合设计中,设计序号为 00329 的突变体分值最低,代表该复合物的自由能最低,则理论上活性也高。因此我们将该突变体命名为 yCas9,并进行实验验证。

### 2.2 Cas9 剪切活性检测

#### 2.2.1 PAM 识别检测

xCas9 可以识别不同的 PAM 序列,包括 NG、GAA 和 GAT。为了验证 Rosetta 设计所得的 yCas9 是否可以识别这些 PAM 序列,我们对其进行了体外 DNA 剪切活性检测(图 2A-C)。由图可见,野生型 SpCas9 主要识别 TGG、TGA 和 TGC,表明其识别的 PAM 的类型有限(图 2A, lanes 2-4)。与文献报道相一致<sup>[9]</sup>,xCas9 可以识别 TGN、GAA 和 GAT 序列(图 2B)。与 xCas9 相比, yCas9 不

表 3 Rosetta 设计结果

Table 3 Results of Rosetta design

ID	Mutation sites							Score
	262	324	409	480	543	694	1 219	
SpCas9	A	R	S	E	E	M	E	
xCas9	T	L	I	K	D	I	V	
00329	A	R	N	K	D	L	T	-1 704.054
00425	A	R	N	K	D	M	T	-1 703.748

仅对含有这些 PAM 序列的 DNA 具有相当的剪切活性 (图 2C), 而且对 GAT 的识别能力比 xCas9 高 (图 2C, lane 7)。由此说明, Rosetta 设计的 yCas9 进一步提升了 xCas9 的 PAM 识别能力。

### 2.2.2 脱靶效应检测

为验证 yCas9 的特异性, 我们通过构建的 8 个错配 sgRNA (表 2, 1-8 号) 和包含 TGG PAM 序列的 DNA 进一步对其进行了体外脱靶检测, 结果如图 3 所示。由图可见, 1-8 号 sgRNA 均可以引导 SpCas9 剪切底物 DNA, 其中, 1-5 号 sgRNA

引导的剪切活性与 0 号 sgRNA 相当 (图 3A, lanes 3-8), 这表明 SpCas9 的脱靶效应比较严重。与 SpCas9 相比, 相同反应条件下, 错配 sgRNA 引导 xCas9 对底物 DNA 的剪切活性均有不同程度的降低 (图 3B, lanes 4-11), 说明 xCas9 的特异性优于 SpCas9。由图 3C 可见, yCas9 的脱靶效应比野生型 SpCas9 低, 但其中 3、4、5、7 号 sgRNA 引导的 yCas9 剪切活性比 xCas9 高 (图 3C, lanes 6、7、8 和 10)。因此, yCas9 的脱靶效应比 xCas9 稍高, 但与野生型 SpCas9 相比大幅度降低。

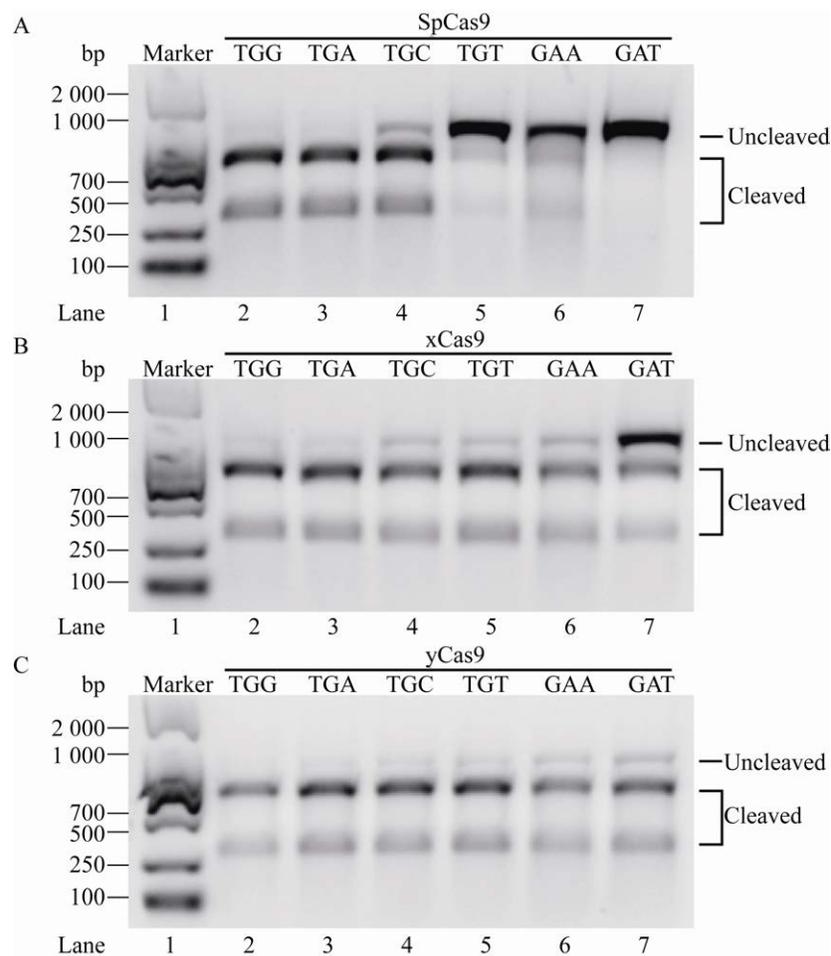


图 2 SpCas9 及其突变体的体外 PAM 识别检测

Fig. 2 *In vitro* PAM compatibility cleavage assay of SpCas9 and its mutants. (A-C) Linearized plasmids containing a target sequence adjacent to TGG, TAG, TGC, TGT, GAA and GAT PAMs were incubated with the Cas9-sgRNA complex (200 nmol/L) at 37 °C for 2 hours. The reaction products were analyzed using an electrophoresis system.

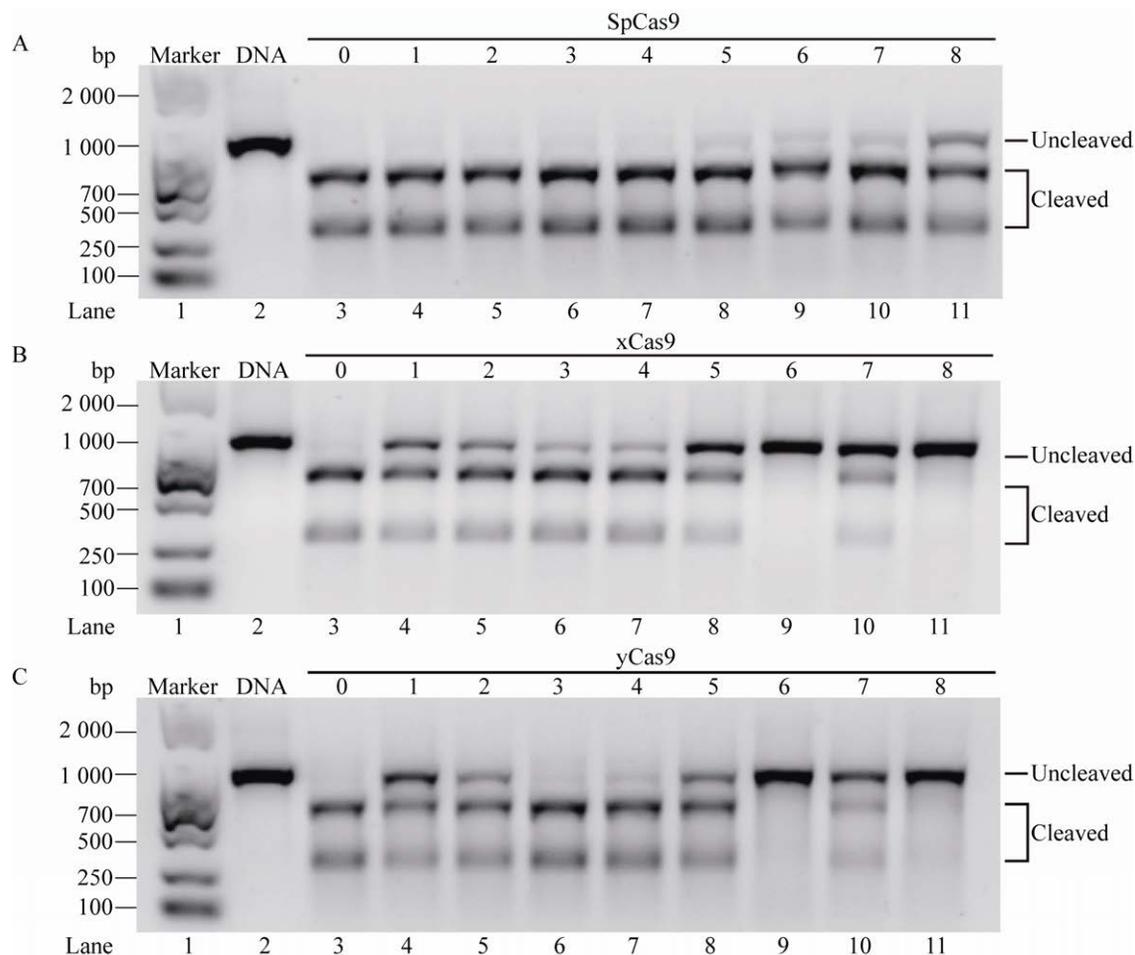


图3 SpCas9 及其突变体的体外脱靶效应检测

Fig. 3 *In vitro* off-target activity assay of SpCas9 and its mutants. (A–C) A linearized plasmid target bearing the TGG PAM was incubated with the Cas9-sgRNA complex (200 nmol/L) at 37 °C for 3 hours. No. 0 sgRNA was perfectly matched and No. 1–8 were partially mismatched. The reaction products were analyzed using an electrophoresis system.

## 2.3 Cas9 剪切机理分析

### 2.3.1 PAM 识别分析

全面了解 Cas9 的 PAM 识别分子机理，有助于我们更好地认识 CRISPR 系统。所以，我们将带有不同 PAM 序列的 Cas9 结构 (TGG: SpCas9, GAT: xCas9 和 yCas9) 进行了 200 ns 的分子动力学模拟。模拟结束后，我们首先对最后一帧构象中 PAM 附近的结构进行单独分析，发现在 xCas9 和 yCas9 中，1 219 位氨基酸的突变破坏了其与 1 335 之间的双齿氢键，增强了 1 333 和 1 335 位氨基酸的结构柔性，从而削弱了周围氨基酸的相互作用 (图 4B 和 C)。然后，我们对 1 219、1 333

和 1 335 位氨基酸的运动轨迹进行了分析。以初始构象为参照，计算其余帧对应构象与初始构象的均方根偏差 (Root mean squared deviation, RMSD)，作时间依赖性的 RMSD 图。由于 MD 模拟的结构在前期没有完全平衡，所以我们选取了 185–200 ns 的模拟轨迹计算 RMSD (图 4D–F)。由图 4D 可知，SpCas9 的 1 333 和 1 335 位氨基酸的 RMSD 值约为  $4 \times 10^{-11}$  m，低于 xCas9 和 yCas9 (图 4E 和 F)。这表明 1 219 位氨基酸突变后增强了 1 333 位和 1 335 位氨基酸结构柔性，进而拓展了 Cas9 蛋白对 PAM 识别能力。以上结果说明，1 219 位氨基酸突变会对 Cas9 的 PAM 识别产生影响。

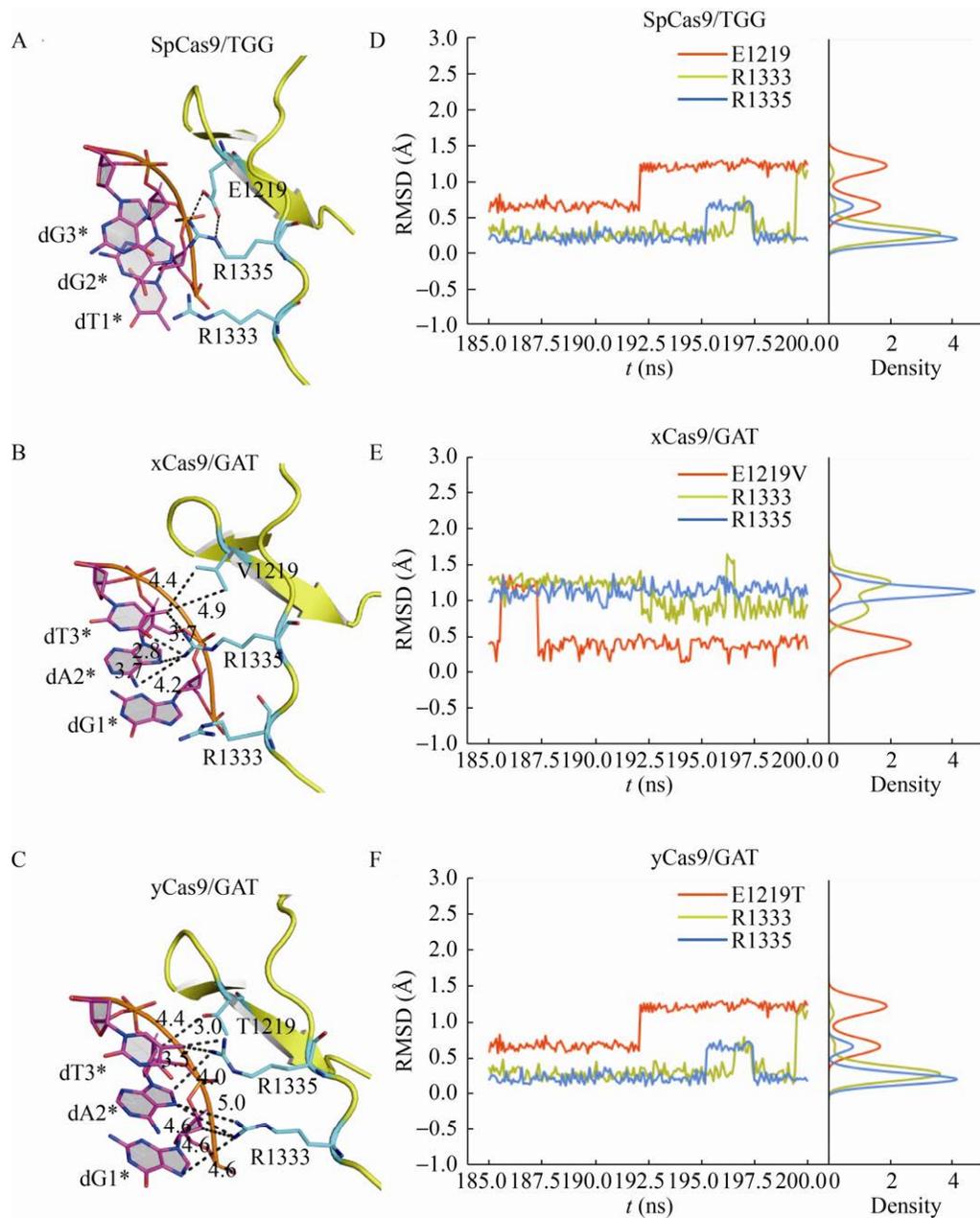


图 4 SpCas9 及其突变体的 PAM 识别图

Fig. 4 Structural insight into PAM recognition by SpCas9 and its mutants. (A–C) Zoom-in views of PAM recognition by SpCas9 (4UN3), xCas9 (6AEG) and yCas9, respectively. (D–F) RMSDs of 1 219, 1 333 and 1 335 residues in SpCas9 (4UN3), xCas9 (6AEG) and yCas9, respectively.

### 2.3.2 脱靶效应分析

另外,为进一步阐明 Cas9 的脱靶机理,本研究构建了带 TGG PAM 序列的 xCas9 和 yCas9 的模拟体系。MD 结束后,我们取出 xCas9 和 yCas9 最后一帧的构象,把它们分别与野生型 SpCas9

的构象叠合,发现这两个突变体的 REC 结构域构象有所改变。因此,分别计算了 xCas9 和 SpCas9、yCas9 和 SpCas9 的 REC 域 RMSD 值(图 5A)。结果显示,xCas9 移动了  $4.13 \times 10^{-11}$  m, yCas9 移动了  $2.47 \times 10^{-11}$  m。最后,我们分析了位于 REC 域

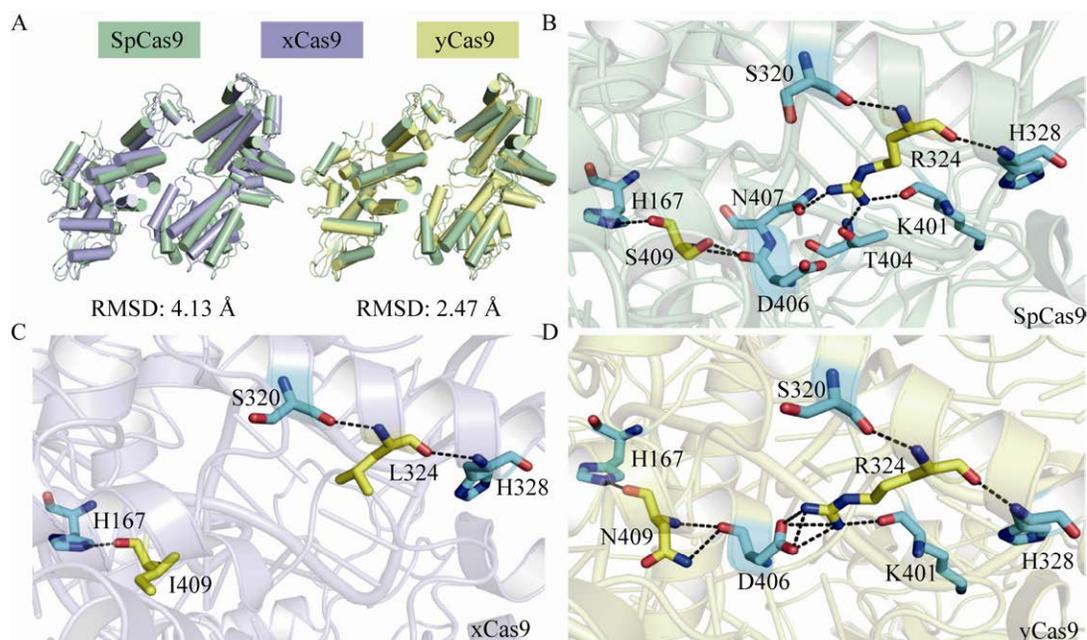


图 5 SpCas9 与其突变体的结构比较

Fig. 5 Structural comparison of SpCas9 with its mutants. (A) Superimposition of the REC lobes of SpCas9 (4UN3), xCas9 (6K4P) and yCas9. (B–D) Zoom-in views of 324 and 409 residues in SpCas9, xCas9 and yCas9, respectively.

的 324 和 409 位的氨基酸 (图 5B–D)。由图可知, 与野生型 SpCas9 相比 (图 5B), xCas9 的 R324 突变成 L324、S409 突变成 I409, 这使得 324 和 409 位的氨基酸与周围氨基酸的氢键作用明显减少, 只与 S320、H328 和 H167 之间存在相互作用 (图 5C)。相同地, 图 5D 表明 yCas9 的这两个氨基酸与周围氨基酸之间的氢键相互作用也相应减少。所以, xCas9 与 yCas9 的 REC 结构域构象的变化影响了 Cas9 脱靶率, 且 REC 域构象改变越大, 脱靶率越低。以上结果说明, Cas9 的 REC 域构象重排会对其特异性产生影响。

### 3 讨论

基因编辑是对生物遗传物质进行特异性改变的一种技术, 常见的有 3 种: 锌指核酸酶技术<sup>[20]</sup>、转录激活因子样效应物核酸酶技术<sup>[21]</sup>和 CRISPR 系统<sup>[22]</sup>。目前, CRISPR 系统应用最为广泛, 主要分为 I 型、II 型和 III 型<sup>[23]</sup>。其中, I 型和 III 型 CRISPR 系统由多个 Cas 蛋白组成, 需要专门的

Cas 核酸内切酶处理 crRNA。相反, II 型系统通过不同的方式处理 crRNA, Cas 蛋白数量明显减少, 所以成为最常用的一种类型。

目前, 已经有多种天然的 II 型 CRISPR 核酸酶用于基因编辑。其中, 来自酿脓链球菌的 Cas9 蛋白 SpCas9 能识别简单的 NGG PAM, 且其在哺乳动物细胞中的剪切活性高, 所以在生物学、生物技术和医学等领域得到了广泛的应用, 例如构建细胞模型、动物模型以及治疗疾病等<sup>[24–25]</sup>。但是, SpCas9 仍然存在许多待解决的挑战性问题, 比如其 PAM 识别范围有限<sup>[10]</sup>、存在脱靶效应等<sup>[3]</sup>。其中, 低兼容性的 PAM 识别模式限制了 SpCas9 在基因编辑领域的应用范围, 而脱靶效应的存在, 则进一步限制了其走向临床应用。为了解决以上问题, 人们对 SpCas9 进行了各种优化改造。

在以往的研究中, 人们主要通过序列比对或者定向进化对 Cas9 蛋白进行改造<sup>[26–28]</sup>。这些方法有一定的盲目性, 往往需要反复试验、耗时耗力。本研究基于 Cas9 蛋白的三维结构, 通过 Rosetta

程序优化 Cas9-DNA 相互作用界面, 以设计筛选突变体。与以往的 SpCas9 改造方法相比, 本文所用的理性设计方法盲目性低、操作简单、效率高, 且具有通用性, 可用于其他 Cas9 蛋白的改造。随着 X 光晶体衍射技术和冷冻电镜技术的快速发展, 越来越多的 Cas9 蛋白结构被解析<sup>[29-31]</sup>, 为该方法的应用提供了坚实的基础。通过分析比较蛋白的电子密度图, 找出对 Cas9 某一功能有重要影响的氨基酸位点, 再结合 Rosetta 进行设计, 以获得新型突变体。因此, 本研究为人们继续开发多 PAM 识别和低脱靶的多功能型 Cas9 蛋白提供了新的途径。

另外, 本文通过上述理性设计方法成功获得了一个 xCas9 新突变体 (yCas9)。体外 PAM 识别检测结果显示, 该突变体具有宽泛的 PAM 识别范围, 可识别 NGG、NGA、NGT、NGC、GAA 和 GAT, 且对 GAT PAM 的识别能力高于 xCas9。脱靶效应检测结果显示, yCas9 的特异性虽然比 xCas9 稍低, 但是与野生型 SpCas9 相比大幅度提高。因此, yCas9 是一个可与 xCas9 媲美的、有潜在应用价值的基因编辑工具。由于本研究的主要目的是验证基于结构的 Cas9 蛋白的优化设计方法是否可行, 因此, 对突变体的活性检测主要在体外进行。当然, 为了更准确地描述突变体在动物细胞内的 PAM 识别能力和脱靶效应, 需要进一步开展体内实验: 关于 PAM 活性检测, 拟将 Cas9 蛋白和 sgRNA 的质粒共转染至含有不同 PAM 序列的荧光报告基因文库的细胞系中进行编辑, 再用流式细胞仪分选出荧光报告基因阳性的细胞, 提取其基因组 DNA, 通过 PCR 构建二代测序文库, 分析并找出产生基因编辑的 PAM 序列<sup>[32]</sup>; 关于脱靶活性检测, 拟在 HEK293T 和 U2OS 细胞中用 GUIDE-seq 对 SpCas9、xCas9 和 yCas9 进行双链断裂的全基因组无偏差鉴定<sup>[25]</sup>。

最后, 本研究还分析了 Cas9 的 PAM 识别和脱靶效应的分子机理, 发现了 E1219 突变为 V/T

后可以提高 R1333 和 R1335 的结构柔性, 从而使 Cas9 可识别较宽泛的 PAM 序列。另外, 我们还发现了 R324 和 S409 突变可以促使 REC 域构象重排, 从而使 Cas9 有高特异性。这些结果为 CRISPR/Cas9 系统的后续改造提供了重要理论指导和坚实基础。

综上所述, 本研究提出了一种用 Rosetta 优化设计 Cas9 蛋白的新方法, 获得了一个多 PAM 识别和低脱靶的新突变体 yCas9, 为生物医学领域提供了有效的基因编辑新工具, 同时也为后续优化 Cas9 蛋白提供了新思路。

## REFERENCES

- [1] Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 2014, 32(4): 347-355.
- [2] Lier C, Baticle E, Horvath P, et al. Analysis of the type II -A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. *Front Genet*, 2015, 6: 214.
- [3] Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 2013, 31(9): 827-832.
- [4] Fu YF, Sander JD, Reyon D, et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*, 2014, 32(3): 279-284.
- [5] Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339(6121): 819-823.
- [6] Kleinstiver BP, Prew MS, Tsai SQ, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 2015, 523(7561): 481-485.
- [7] Chen JS, Dagdas YS, Kleinstiver BP, et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, 2017, 550(7676): 407-410.
- [8] Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 2016, 529(7587): 490-495.
- [9] Hu JH, Miller SM, Geurts MH, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, 2018, 556(7699): 57-63.

- [10] Nishimasu H, Shi X, Ishiguro S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, 2018, 361(6408): 1259-1262.
- [11] Legut M, Daniloski Z, Xue XH, et al. High-throughput screens of PAM-flexible Cas9 variants for gene knockout and transcriptional modulation. *Cell Rep*, 2020, 30(9): 2859-2868.e5.
- [12] Kuhlman B, Dantas G, Ireton GC, et al. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 2003, 302(5649): 1364-1368.
- [13] Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol*, 2011, 487: 545-574.
- [14] Broo KS, Brive L, Ahlberg P, et al. Catalysis of hydrolysis and transesterification reactions of *p*-nitrophenyl esters by a designed helix-loop-helix dimer. *J Am Chem Soc*, 1997, 119(47): 11362-11372.
- [15] Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 2015, 1-2: 19-25.
- [16] Ray A, Di Felice R. Protein-mutation-induced conformational changes of the DNA and nuclease domain in CRISPR/Cas9 systems by molecular dynamics simulations. *J Phys Chem B*, 2020, 124(11): 2168-2179.
- [17] Govindarajan S, Goldstein RA. On the thermodynamic hypothesis of protein folding. *Proc Natl Acad Sci USA*, 1998, 95(10): 5545-5549.
- [18] Anfinsen CB. Principles that govern the folding of protein chains. *Science*, 1973, 181(4096): 223-230.
- [19] Alford RF, Leaver-Fay A, Jeliazkov JR, et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput*, 2017, 13(6): 3031-3048.
- [20] Urnov FD, Rebar EJ, Holmes MC, et al. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 2010, 11(9): 636-646.
- [21] Miller JC, Tan SY, Qiao GJ, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, 2011, 29(2): 143-148.
- [22] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007, 315(5819): 1709-1712.
- [23] Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*, 2011, 9(6): 467-477.
- [24] Xue W, Chen SD, Yin H, et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*, 2014, 514(7522): 380-384.
- [25] Matano M, Date S, Shimokawa M, et al. Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat Med*, 2015, 21(3): 256-262.
- [26] Casini A, Olivieri M, Petris G, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol*, 2018, 36(3): 265-271.
- [27] Slaymaker IM, Gao LY, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 2016, 351(6268): 84-88.
- [28] Ma DC, Xu ZM, Zhang ZY, et al. Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information. *Nat Commun*, 2019, 10: 560.
- [29] Nishimasu H, Cong L, Yan WX, et al. Crystal Structure of *Staphylococcus aureus* Cas9. *Cell*, 2015, 162(5): 1113-1126.
- [30] Huai C, Li G, Yao RJ, et al. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat Commun*, 2017, 8: 1375.
- [31] Yamano T, Nishimasu H, Zetsche B, et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*, 2016, 165(4): 949-962.
- [32] Hu ZY, Wang S, Zhang CD, et al. A compact Cas9 ortholog from *Staphylococcus auricularis* (SauriCas9) expands the DNA targeting scope. *PLoS Biol*, 2020, 18(3): e3000686.

(本文责编 郝丽芳)