

(SPRUC 2014 Young Scientist Award 受賞 研究報告)

CRISPR-Cas ヌクレアーゼの結晶構造

東京大学大学院 理学系研究科
西増 弘志

Abstract

原核生物のもつ CRISPR-Cas 系は外来核酸に対する獲得免疫機構としてはたらく。CRISPR-Cas 系にかかわる RNA 依存性ヌクレアーゼである Cas9 や Cpf1 はガイド RNA と相補的な二本鎖 DNA を特異的に切断する。近年、この性質を応用したゲノム編集技術の登場により、様々な生物のゲノム情報（生命の設計図）を「書き換える」ことが可能になってきた。これまでに我々は Cas9 および Cpf1 の結晶構造を決定し、その RNA 依存的な DNA 切断機構を解明してきた。また、Cas9 と Cpf1 の構造比較から、CRISPR-Cas ヌクレアーゼの作動機構の共通性および多様性が明らかになった。さらに、これらの構造情報は新規のゲノム改変ツールの開発にも大きく貢献してきた。

1. CRISPR-Cas 系

細菌などの原核生物のもつ CRISPR-Cas (clustered regularly interspaced short palindromic repeat-CRISPR-associated) 系はウイルスなどの外来核酸からの防御を担う獲得免疫機構としてはたらく^[1]。CRISPR-Cas 系は、(1) Adaptation、(2) Expression、(3) Interference の3つのステップからなり、crRNA (CRISPR RNA) と複数の Cas タンパク質が関与する。Interference のステップにおいて、特定の Cas タンパク質と crRNA が Cas-crRNA 複合体を形成し、crRNA と相補的な外来核酸を認識・切断する。Cas-crRNA 複合体の構造にもとづき、CRISPR-Cas 系は2つのクラスに分類される^[2,3]。クラス1 CRISPR-Cas 系には複数の Cas タンパク質からなる複合体が関与する一方、クラス2 CRISPR-Cas 系には単一の Cas タンパク質からなる複合体が関与する。クラス2 CRISPR-Cas 系はII型、V型、VI型に分類される。

II型 CRISPR-Cas 系においては Cas9 タンパク質が2種類のガイド RNA (crRNA および tracrRNA (trans-activating crRNA)) と Cas9-crRNA-tracrRNA 複合体を形成し標的二本鎖 DNA を切断する^[4] (図1A)。Cas9 は2つのヌクレアーゼドメイン (RuvC/HNH) をもち、HNH ドメインは標的 DNA のうちガイド RNA と相補的な DNA 鎖 (相補鎖) を切断する一方、RuvC

ドメインはもう一方の DNA 鎖 (非相補鎖) を切断する。標的 DNA の認識にはガイド RNA との相補性に加え、PAM (protospacer adjacent motif) とよばれる特定の塩基配列が標的配列の近傍に存在する必要がある。異なる生物種に由来する Cas9 のアミノ酸配列は多様であり、認識するガイド RNA や PAM の配列が異なる。たとえば、*Streptococcus pyogenes* 由来 Cas9 (SpCas9)^[4]、*Staphylococcus aureus* 由来 Cas9 (SaCas9)^[5]、*Francisella novicida* 由来 Cas9 (FnCas9)^[6] はそれぞれ NGG (N は任意の塩基)、NNGRRT (R は A または G)、NGG という配列を PAM として認識する。ガイド RNA のガイド配列 (20 塩基) は変更可能であるため、Cas9-sgRNA 複合体は PAM をもち、ガイド配列と相補的な標的二本鎖 DNA を特異的に

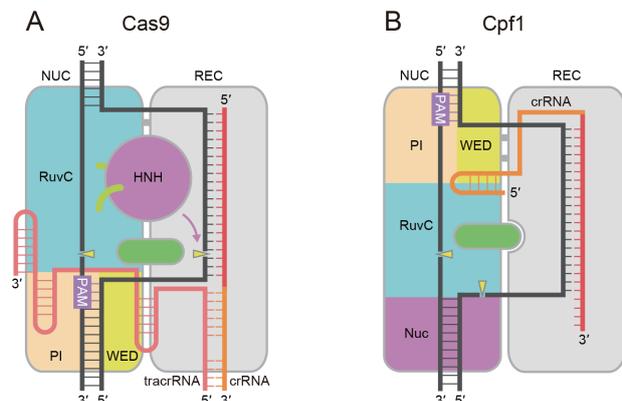


図1 Cas9 と Cpf1 の作動機構

切断することができる。さらに、crRNA と tracrRNA を連結した sgRNA (single-guide RNA) も同様に機能する^[4]。したがって、Cas9-sgRNA は効率的なゲノム編集ツールとして急速に普及してきた^[7]。さらに、DNA 切断活性をもたない Cas9 変異体 (dCas9) は RNA 依存的にゲノム DNA の狙った位置に結合するため、この性質を利用した様々な新規技術も開発されている。

V 型 CRISPR-Cas 系においては Cpf1 (CRISPR from *Prevotella* and *Francisella*) が crRNA と複合体を形成し標的二本鎖 DNA を切断する^[8] (図 1B)。Cpf1 は、(1) crRNA のみを利用し tracrRNA を必要としない、(2) TTTV (V は T 以外の塩基) という配列を PAM として認識する、(3) PAM から離れた位置で標的 DNA を切断し突出末端をつくる、(4) RuvC ドメインをもつが HNH ドメインをもたない、など Cas9 と異なる特徴をもつ。哺乳類細胞において活性を示す *Acidaminococcus sp.* 由来 Cpf1 (AsCpf1) や *Lachnospiraceae bacterium* 由来 Cpf1 (LbCpf1) は、Cas9 と異なる特性をもつゲノム編集ツールとして注目されている。

2. SpCas9 の結晶構造

2012 年、Cas9 は新規の RNA 依存性ヌクレアーゼであることが報告されたが^[4]、その作動機構は不明であった。そこで、Cas9 による DNA 切断機構の解明を目指し、SpCas9-sgRNA-DNA 複合体の結晶構造を決定した^[9] (図 2A)。結晶構造から、SpCas9 は 2 つのローブから構成されることが明らかになった。一方のローブは α ヘリックスからなる新規フォールドをもち、sgRNA-DNA の認識に関与していた。そこで REC (recognition) ローブと命名した。もう一

方のローブは 2 つのヌクレアーゼドメイン (RuvC/HNH) と C 末端ドメインから構成されていた。そこで NUC (nuclease) ローブと命名した。C 末端ドメインは新規フォールドをもち PAM と相互作用するのに適した位置に存在していた。構造情報をもとにした変異体解析の結果から、C 末端ドメインは PAM の認識にかかわることが示唆された。そこで C 末端ドメインを PI (PAM-interacting) ドメインと命名した。後日報告された SpCas9-sgRNA-DNA (PAM を含む) 複合体の結晶構造から、PI ドメインの Arg1333 と Arg1335 が NGG PAM を認識していることが明らかになった^[10]。RuvC ドメインは非相補鎖 DNA を切断するのに適した位置に存在していた一方、HNH ドメインは相補鎖 DNA と離れた位置に存在していた。したがって、DNA 切断の際、HNH ドメインは構造変化を起こすことが示唆された。sgRNA のガイド配列は相補鎖 DNA と RNA-DNA ヘテロ二本鎖を形成し、2 つのローブの間に結合していた。一方、sgRNA の他の領域は特徴的な立体構造をとり、Cas9 と広範囲に相互作用し複合体形成に貢献していた。以上の結果から、Cas9 の RNA 依存性 DNA 切断機構が説明された。

SpCas9-sgRNA-DNA 複合体の結晶構造から、sgRNA の 2 つのループ領域は溶媒側に露出していることが明らかになった^[9]。この 2 つのループ領域に MS2 ファージ由来コートタンパク質を特異的に認識する RNA 配列を融合した改変型 sgRNA を作製し、(1) dCas9、(2) 改変型 sgRNA、および、(3) 転写活性化因子を融合した MS2 コートタンパク質を哺乳類細胞に共発現させることにより、sgRNA 依存的に標的遺伝子を活性化することに成功した^[11]。

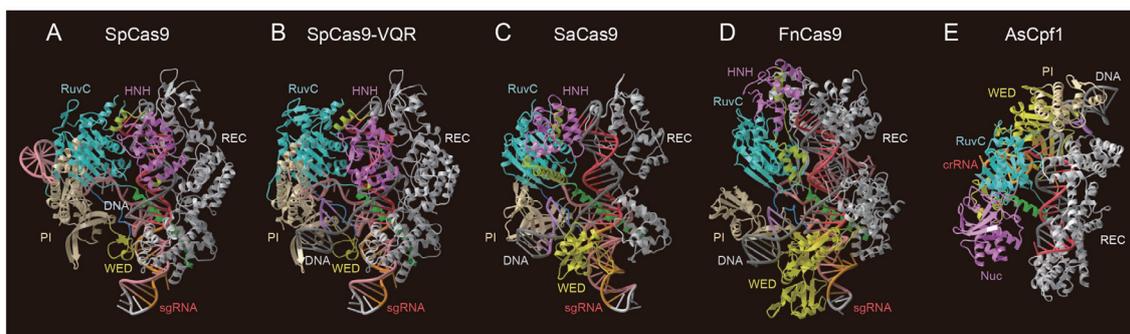


図 2 Cas9 と Cpf1 の結晶構造

(A) SpCas9-sgRNA-DNA 複合体

(B) VQR 改変型 -sgRNA-DNA 複合体

(C) SaCas9-sgRNA-DNA 複合体

(D) FnCas9-sgRNA-DNA 複合体

(E) AsCpf1-crRNA-DNA 複合体

3. SpCas9 改変体の結晶構造

SpCas9-sgRNA-DNA 複合体の構造情報は、DNA 切断精度の高い SpCas9 改変体^[12,13] や異なる PAM 配列を認識する SpCas9 改変体^[14] の開発にも貢献してきた。野生型 SpCas9 は NGG PAM を認識する一方、VQR 改変体 (D1135V/R1335Q/T1337R)、EQR 改変体 (D1135E/R1335Q/T1337R)、VRER 改変体 (D1135V/G1218R/R1335E/T1337R) はそれぞれ NGA、NGAG、NGCG を PAM として認識する^[14]。SpCas9 改変体による PAM 認識機構を解明するため、3種類の SpCas9 改変体に関して、SpCas9-sgRNA-DNA 複合体の結晶構造を決定した^[15] (図 2B)。野生型 SpCas9 において PAM の3文字目の G は Arg1335 によって認識される一方、VQR/EQR 改変体において PAM の3文字目の A は R1335Q によって認識されていた。VRER 改変体においては PAM の3文字目の C は R1335E によって認識されていた。野生型 SpCas9 と SpCas9 改変体の構造比較から、R1335Q/R1335E 以外のアミノ酸置換は、PAM 近傍の二本鎖 DNA の糖-リン酸骨格の構造変化を誘起し、PAM の3文字目の塩基と R1335Q/R1335E との間の直接的な水素結合を可能にしていることが明らかになった。

4. SaCas9 の結晶構造

SaCas9 (1053 残基) は SpCas9 (1368 残基) に比べて小型であるため、ウイルスベクターへの導入効率の高いゲノム編集ツールとして注目されている^[5]。SaCas9 の作動機構の解明を目指し、SaCas9-sgRNA-DNA 複合体の結晶構造を決定した^[16] (図 2C)。SaCas9 は SpCas9 と同様に、2つのローブからなる構造をもち、RNA-DNA ヘテロ二本鎖は2つのローブの間に結合していた。したがって、RNA 依存的な DNA 認識機構は両者において高度に保存されていることが示唆された。一方、SpCas9 と SaCas9 の PAM 特異性の違いと一致して、SaCas9 は PI ドメインの Asn985、Asn986、Arg991、Arg1015 を用いて NNGRRT PAM を認識していた。SpCas9 と SaCas9 の構造比較から、両者の間のドメイン構造の違いも明らかになった。SpCas9 と比較し、SaCas9 はコンパクトな REC ローブをもっていた。さらに、SpCas9 と SaCas9 において二本鎖 DNA とガイド RNA の間に結合するドメイン (WED ドメイン) は大きく構造が異なっていた。これらの比較から、SpCas9 と SaCas9 は構造の異なる

REC ドメインおよび WED ドメインを用いて sgRNA の構造の違いを特異的に認識していることが明らかになった。さらに、構造情報をもとに、SaCas9 を用いた転写活性化ツールおよび、誘導型 SaCas9 の作製に成功した^[16]。

5. FnCas9 の結晶構造

FnCas9 (1629 残基) は Cas9 タンパク質の中でも最もサイズが大きく、SpCas9 や SaCas9 との配列相同性も低い。FnCas9 の作動機構を解明するために、FnCas9-sgRNA-DNA 複合体の結晶構造を決定した^[6] (図 2D)。FnCas9 は SpCas9 や SaCas9 と類似の RuvC/HNH ドメインをもつ一方、FnCas9 の REC/WED ドメインは新規フォールドをとっていた。REC/WED ドメインの構造の違いと一致して、sgRNA の構造も大きく異なっていた。結晶構造から、NGG PAM は PI ドメインの Arg1556、Arg1585 によって認識されることが明らかになった。SpCas9 と FnCas9 の構造比較から、両者の NGG PAM 認識機構は異なることが明らかになった。さらに、構造情報をもとに FnCas9 に3つの変異を導入し、YG (Y は T または C) という配列を PAM として認識する FnCas9 改変体の作製に成功した^[6]。

6. AsCpf1 の結晶構造

Cpf1 は RuvC ドメインを除き Cas9 を含む既知タンパク質と相同性をもたないため、その DNA 切断機構は不明だった。Cpf1 による DNA 切断機構を解明するために、AsCpf1-crRNA-DNA 複合体の結晶構造を決定した^[17] (図 2E)。結晶構造から、Cpf1 は2つのローブ (REC/NUC) からなることが明らかになった。REC ローブは2つのドメイン (REC1/REC2) から構成され、NUC ローブは4つのドメイン (RuvC/WED/PI/Nuc) から構成されていた。RuvC ドメインを除く5つのドメインは新規フォールドをもっていた。crRNA の5' 末端領域はシュドノット構造をとり WED/RuvC ドメインによって認識されていた。一方、crRNA のガイド配列は標的 DNA と RNA-DNA ヘテロ二本鎖を形成し、2つのローブの間に収容されていた。TTTA PAM を含む二本鎖 DNA は歪んだ二重らせん構造をとり、WED/REC1/PI ドメインにより構造・配列特異的に認識されていた。Nuc ドメインは RuvC ドメインの近傍に位置し、相補鎖 DNA の切断に関与することが明らかになった。RuvC ドメインを除き Cas9 と

Cpf1は配列相同性をもたないが、両者は2つのローブから構成される全体構造を共通してもっていた。一方、PAM認識・DNA切断機構は大きく異なっていた。これらの比較から、Cas9とCpf1の間の機能収斂が明らかになった。

7. おわりに

近年の構造機能研究の進展により、Cas9やCpf1によるDNA切断機構の大枠は解明された。しかし、DNA切断におけるダイナミクスや触媒反応の詳細は不明である。さらに、最近の研究から、新規のCRISPR-Casヌクレアーゼも発見されている^[18,19]。今後の研究により、多様なCRISPR-Casヌクレアーゼの作動機構の全容解明、および、新規のゲノム改変ツールの開発が期待される。

謝辞

本研究は、MITのFeng Zhang博士、東京大学大学院理学系研究科の石谷隆一郎准教授、濡木理教授との共同研究として行った。

X線回折実験は、SPRING-8のBL32XUおよびBL41XU(課題番号2014A1356、2014B1223、2015A0119、2015B0119、2016A0119、2016B0119)において行った。

本研究は、JST戦略的創造研究推進事業(さきがけ)、JSPS科研費(26291010、15H01463)、文部科学省及び国立研究開発法人日本医療研究開発機構創薬等ライフサイエンス研究支援基盤事業(創薬等支援技術基盤プラットフォーム事業)の支援を受けて行った。

参考文献

- [1] R. Barrangou *et al.*: *Science* **315** (2007) 1709-1712.
- [2] K. S. Makarova *et al.*: *Nat Rev Microbiol* **13** (2015) 722-736.
- [3] H. Nishimasu and O. Nureki: *Curr Opin Struct Biol* **43** (2016) 68-78.
- [4] M. Jinek *et al.*: *Science* **337** (2012) 816-821.
- [5] F. A. Ran *et al.*: *Nature* **520** (2015) 186-191.
- [6] H. Hirano *et al.*: *Cell* **164** (2016) 950-961.
- [7] L. Cong *et al.*: *Science* **339** (2013) 819-823.
- [8] B. Zetsche *et al.*: *Cell* **163** (2015) 759-771.
- [9] H. Nishimasu *et al.*: *Cell* **156** (2014) 935-949.

- [10] C. Anders, O. Niewoehner, A. Duerst and M. Jinek: *Nature* **513** (2014) 569-573.
- [11] S. Konermann *et al.*: *Nature* **517** (2015) 583-588.
- [12] B. P. Kleinstiver *et al.*: *Nature* **529** (2016) 490-495.
- [13] I. M. Slaymaker *et al.*: *Science* **351** (2016) 84-88.
- [14] B. P. Kleinstiver *et al.*: *Nature* **523** (2015) 481-485.
- [15] S. Hirano, H. Nishimasu, R. Ishitani and O. Nureki: *Mol Cell* **61** (2016) 886-894.
- [16] H. Nishimasu *et al.*: *Cell* **162** (2015) 1113-1126.
- [17] T. Yamano *et al.*: *Cell* **165** (2016) 949-962.
- [18] O. O. Abudayyeh *et al.*: *Science* **353** (2016) aaf5573.
- [19] D. Burstein *et al.*: *Nature* doi:10.1038/nature21059 (2016).

※用語解説

・ゲノム編集

細胞の中でゲノムDNAが切断されると、修復機構により切断部位の再結合が起こる。この際、塩基の欠損や挿入などの変異が偶発的に生じる。ゲノム編集技術は、ゲノムDNAの狙った部位を人為的に切断し、修復過程における偶発的な変異を利用することによりゲノム情報を「書き換える」技術である。制限酵素は6塩基程度を認識し二本鎖DNAを切断するのに対し、ZFN(zinc finger nuclease)やTALEN(transcription activator-like effector nuclease)などの人工ヌクレアーゼは20塩基以上の配列を認識し二本鎖DNAを切断する。したがって、人工ヌクレアーゼを用いることによりゲノムDNAの標的部位を選択的に切断しゲノム編集を行うことが可能である。しかし、ZFNやTALENはタンパク質モジュールにより塩基配列を認識するため、標的配列ごとにタンパク質を設計しなければならないという問題点があった。一方、Cas9はsgRNAのもつガイド配列(20塩基)と相補的な二本鎖DNAを切断し、sgRNAのガイド配列は変更することができるため、簡便で迅速なゲノム編集ツールとして広く普及した。

西増 弘志 NISHIMASU Hiroshi

東京大学大学院 理学系研究科
〒113-0032 東京都文京区弥生2-11-16
TEL : 03-5841-4391
e-mail : nisimasu@bs.s.u-tokyo.ac.jp