Published in final edited form as:

Methods Enzymol. 2019 ; 616: 27–41. doi:10.1016/bs.mie.2018.11.003.

Reconstitution and biochemical characterization of ribonucleoprotein complexes in Type I-E CRISPR-Cas systems

Yibei Xiao¹ and Ailong Ke^{2,*}

¹State Key Laboratory of Natural Medicines, Department of Pharmacology, China Pharmaceutical University, Nanjing 210009, P. R. China.

²Department of Molecular Biology and Genetics, Cornell University, 253 Biotechnology Building, Ithaca, NY 14853, USA.

Abstract

Type I CRISPR-Cas, the most prevalent CRISPR system, features a sequential target searching and degradation process. First, the multi-subunit surveillance complex Cascade recognizes the matching dsDNA target flanked by protospacer adjacent motif (PAM), promotes the heteroduplex formation between CRISPR RNA (crRNA) and the target strand (TS) DNA and displaces the non-target strand (NTS) DNA, resulting in R-loop formation. The helicase-nuclease fusion enzyme Cas3 is then specifically recruited to Cascade/R-loop, nicks and processively degrades the DNA target. Here, by using type I-E CRISPR-Cas system from *Thermobifida fusca*, we provide protocols for the biochemical reconstitution of the Cascade/R-loop and Cascade/R-loop/Cas3 complexes that allowed for high resolution structure determination and mechanism investigation. The procedures may be applicable for structural and mechanistic investigations of other type I CRISPR-Cas systems, and may serve as a reference for the study of other multi-component protein-nucleic acid complexes.

Keywords

CRISPR-Cas; Cascade; R-loop; Cas3

1 Introduction

CRISPR-Cas is an RNA-based adaptive immune system found in bacteria and archaea. It consists of a clustered regularly interspaced short palindromic repeats (CRISPR) array and an associated operon (*cas*). It protects prokaryotes against foreign genetic elements such as phages and plasmids through RNA-guided nucleic acid degradation (Barrangou et al., 2007; Brouns et al., 2008; Hale et al., 2009; Makarova et al., 2006; Marraffini and Sontheimer, 2008; Mojica et al., 2005; Pourcel et al., 2005). Based on the unique features in the *cas* operon, the CRISPR-Cas systems can be divided into two classes and at least six types therein (Makarova et al., 2015; Shmakov et al., 2015). All Class 1 systems (Type I, III, and

*Correspondence to ailong.ke@cornell.edu.

IV) employ a multi-subunit complex, whereas Class 2 systems (Type II, V, and VI) use a single-protein effector to achieve CRISPR interference.

Type I systems are the most prevalent CRISPR system. Among the seven subtypes within Type I (I-A to I-G), Type I-E CRISPR is the most well studied (Makarova and Koonin, 2015). In type I CRISPR interference, the CRISPR-Associated Complex for Antiviral Defense (Cascade) uses crRNA guide to identify 32–35 bp of matching dsDNA (protospacer) flanked by an optimal Protospacer Adjacent Motif (PAM) (Mojica et al., 2009). Following DNA unwinding, duplex formation between crRNA spacer and target DNA strand, and displacement of the non-target DNA strand, the nascent R-loop becomes stably locked by the Cascade (Brouns et al., 2008; Rutkauskas et al., 2015; Westra et al., 2012; Wiedenheft et al., 2011). Subsequently, the nuclease-helicase enzyme Cas3 is recruited to processively degrade the non-target and target DNA strands (Hochstrasser et al., 2014; Mulepati and Bailey, 2013; Sinkunas et al., 2013; Westra et al., 2012).

Structural insight on Cascade was first derived from the *E. coli* Type I-E system (Jackson et al., 2014; Mulepati et al., 2014; Wiedenheft et al., 2011; Zhao et al., 2014). *Eco*Cascade was shown to adopt a seahorse-shaped architecture. The structure can be roughly divided into a helical backbone, composed from head to tail of Cas5e, Cas7_{1–6}, and the crRNA processing enzyme Cas6e, and an inner belly composed of Cse1 and two Cse2 proteins. The former section is responsible for displaying the crRNA and base-pairing with the target, and the latter contains the putative functions of PAM recognition and Cas3 recruitment. The specific mechanism for PAM recognition was later elucidated using an *Eco*Cascade/R-loop mimic structure (Hayes et al., 2016). However, further biochemical reconstitution and structural analysis of Cascade/R-loop/Cas3 complex was hindered by the poorly behaving *E. coli* Cas3 protein (Hochstrasser et al., 2014). To overcome the technical challenge, we screened multiple Cas3 homologs and identified a well-behaving Cas3 from the *Thermobifida fusca* Type I-E system (Huo et al., 2014). Interestingly, the *E. coli* Cascade and *T. fusca* Cas3 do not hetero-assemble *in vitro*, suggesting that the molecular recognition between DNA targeting and degradation machinery is very stringent, possibly due to the evolutionary pressure to prevent cross-talks between different Type I-E CRISPR systems. We therefore derived protocols to recombinantly assemble the *T. fusca* Cascade from the *E. coli* cells, and reconstituted the entire CRISPR interference reaction *in vivo*. Thorough biochemical investigation of the complex enabled us to synchronize the *T. fusca* Cascade into different stages of the R-loop formation process, and to further program *Tfu*Cascade/R-loop into the Cas3-bound state. These efforts in turn led to a series of high-resolution cryo-electron microscopy (EM) snapshots that elucidated the key mechanisms of the RNA-guided DNA degradation process in Type I-E systems (Dillard et al., 2018; Huo et al., 2014; Jung et al., 2017; Xiao et al., 2017; Xiao et al., 2018).

In this chapter, we present the biochemical pipeline and relevant considerations that lead to the successful reconstitution of these functional ribonucleoprotein complexes in type I-E system. PAM recognition is an essential step that distinguishes CRISPR-mediated DNA recognition from RNAi-mediated RNA recognition. To identify the strongest PAM sequence, we initially used an *in vivo* assay to quantify target plasmid loss. This assay was also used to evaluate Cascade and Cas3 mutants *in vivo*. Electrophoretic mobility shift assays (EMSAs)

are an important tool in several steps of our study. We used them to probe target DNA binding and to evaluate the quality of the assembled ribonucleoprotein complexes. In addition, chemical probing was used to monitor the R-loop formation process. Affinity purification and size exclusion chromatography (SEC) were employed to assemble high purity ribonucleoprotein complexes. These assays are applicable to other multicomponent ribonucleoprotein assemblies in suitable scenarios.

***In vivo* assay for Cascade and Cas3-mediated target plasmid loss.**

Since the natural PAM sequences of the type I-E CRISPR-Cas system from *Thermobifida fusca* are unknown, we first set up an *in vivo* CRISPR interference assay in *Escherichia coli* to screen for PAM sequences that lead to efficient CRISPR interference by the *T. fusca* Type I-E system. Because *T. fusca* itself is not amenable to genetic manipulation, we cloned components of the I-E CRISPR system into plasmids and recombinantly assembled the *T. fusca* system in *E. coli*. To avoid possible interference from the native *E. coli* Type I-E system, we choose the *E. coli* BL21-AI cell line, in which the native *cas* operon has been knocked out. This cell line has the added benefit that the T7 RNA polymerase is under tight control by the arabinose promoter, preventing the leaky expression of CRISPR components that often resulted in cell toxicity. We divided the *T. fusca* CRISPR-Cas operon into several plasmids in order to recombinantly express the *Tfu*Cascade and *Tfu*Cas3 in *E. coli*. The *Tfu*Cas3 gene was inserted into a pET28b (Kan^R) vector. The *Tfu*Cascade expression construct was generated by inserting the *T. fusca* Cse1/Cse2/Cas7/Cas5/Cas6e PCR-fragment into a pBAD vector (Ap^R). The pre-crRNA expression cassette containing four identical CRISPR units under the control of the T7 RNA polymerase promoter and terminator was synthesized by GenScript and cloned into the pACYC-Duet-1 vector(Cm^R). To generate target DNA with variant PAM sequences, oligo pairs containing a 32-bp matching protospacer and a variety of 3-bp PAM sequences were annealed, generating the sticky ends of BamHI and XhoI restriction sites. The annealed primer pairs were phosphorylated at the 5'-termini by T4 PNK, and ligated into a linearized pCDFDuet-1 (Sm^R) vector. The pET28b_Cas3, pBAD-Cascade, pACYC-CRISPR and pCDF-target plasmids were transformed into BL21-AI competent cells and grown on LB plates containing kanamycin (50 µg/ml), ampicillin (100 µg/ml), streptomycin (50 µg/ml) and chloramphenicol (34 µg/ml). Upon induction, if the expressed *Tfu*Cascade and *Tfu*Cas3 degrade target plasmids containing the protospacer flanked by an interference PAM, the cells will not grow on agar plates containing streptomycin. To quantify interference efficacy, the number of obtained colonies was counted, and the relative interference efficiency was calculated as the ratio of colonies on the nonselective over selective plates. PAM sequence evaluated this way exhibited 5'-WAK consensus sequences, with 5'-AAG being the most potent.

The following protocol describes the detailed procedure for the plasmid loss assay.

1. Co-transform the four plasmids into BL21-AI cells and grow on LB plates containing kanamycin (50 µg/ml), ampicillin (100 µg/ml), streptomycin (50 µg/ml) and chloramphenicol (34 µg/ml).

2. Pick single colonies, culture at 37 °C in 1 mL LB medium containing all four antibiotics for 6 hours.
3. Transfer 10 µL culture into 1 mL nonselective LB medium, culture at 37 °C for around one and a half hours until O.D.₆₀₀ reaches 0.3.
4. Add 0.5% L-arabinose and 2.5 mM IPTG to induce the expression of *TfuCascade*, Cas3, and pre-crRNA. Induce expression at 37 °C for 12 hours. A shorter induction time will lead to slightly reduced dynamic range in CRISPR interference efficiency.
5. Make a series of ten-fold dilutions for each cell culture. Plate 100 µL of each sample onto Kan + Ap + Cm LB plates (nonselective for pCDF target) and Kan + Ap + Cm + Sm plates (selective for the target plasmid). Count the number of colonies on each plate after overnight incubation at 37 °C. The CRISPR interference efficiency is determined as the ratio of colony-forming units on the nonselective over selective plates. Ten-fold titration series were routinely plated to capture the dynamic range. Each experiment was repeated three times to calculate the standard deviation.

The target plasmid loss assay can be adapted with minor modifications to evaluate the functional importance of residues and structural features in Cascade and Cas3. First, the target plasmid containing the strongest PAM sequence, 5'-AAG, was used. Second, the expression was induced at 37 °C for only 6 hours before plating. The shorter induction time would result in a slightly reduced dynamic range in CRISPR interference efficiency. However, this disadvantage is outweighed by a substantially reduced error spread, thus making the mutagenesis data more reliable for quantitative comparison.

Cloning, expression, and purification of the *T. fusca* Cascade complex

In previous studies, *EcoCascade* was assembled by co-expressing the *cse1* cassette, the *cse2-cas7-cas5e-cas6e* gene cassette, and the pre-crRNA expression cassette from three separate plasmids in *E. coli* cells, with subsequent purification of the complex via the His₆-tag at the N-terminal of *cse2*. Applying the same strategy to *TfuCascade*, we were only able to obtain the Cse2-Cas7-Cas5e-Cas6e-crRNA sub-complex; Cse1 was severely sub-stoichiometric. To overcome the stoichiometry problem, we moved the affinity tag from Cse2 to Cse1. The optimized cloning, expression and purification protocol is as follows. The *cse1* gene was cloned into a modified Twin-Strep-SUMO-pET19b vector (Amp^R). The *cse2-cas7-cas5e-cas6e* gene cassette was cloned into a pCDFDuet-1 vector (Str^R). A synthetic pre-crRNA expression cassette containing 4 identical copies of the repeat-spacer sequences was cloned into the pRSFDuet-1 vector (Kan^R). All three sequence-verified plasmids were co-transformed into the *E. coli* BL21 (DE3) Star cells. The cell culture was grown in LB medium at 37°C until the optical density at 600 nm reached 0.8. Expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM at 25°C overnight. Cells were harvested by centrifugation and lysed by sonication in buffer A containing 50 mM HEPES pH 7.5, and 500 mM NaCl. The lysate was centrifuged at 15,000 rpm for 60 min at 4°C, and the supernatant was applied onto pre-equilibrated Strep-Tactin Superflow resin (IBA, Göttingen, Germany). After washing with 100 mL of buffer A. The

protein was eluted with buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, and 5 mM desthiobiotin), and the eluted proteins were incubated with SUMO-protease at 4°C overnight. The *Tfu*Cascade was further purified with size-exclusion chromatography (SEC, HiLoad 16/60 Superdex 200 hg; GE Healthcare) equilibrated with buffer C (10 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT). The peak fractions were pooled and further purified with a second round of size-exclusion chromatography equilibrated with buffer C. The peak fractions were then concentrated to 10 mg/mL, and snap-frozen in liquid nitrogen and stored at -80°C for later usage. *Tfu*Cascade purified this way exhibited good stoichiometry, and has a 260/280 ratio of ~1.07.

Electrophoretic Mobility Shift Assay

EMSA (Electrophoretic Mobility Shift Assay) is a widely used technique to study protein-DNA or protein-RNA interactions. This technique relies on the principle that protein-nucleic acid complexes migrate slower relative to unbound nucleic acid during electrophoresis in a non-denaturing polyacrylamide (PAGE) or agarose gel. The extent of retardation is dependent on the mass, shape, and overall electrostatic charge of the protein-nucleic acid complex. EMSA can be utilized to identify sequence-specificity of nucleic acids-binding proteins, or to quantitatively measure the thermodynamic and kinetic parameters of the binding reaction. Usually, the concentration of the nucleic acid substrate is kept as low as practically possible, typically at least 10-fold lower than the anticipated binding dissociation constant, to avoid sequence-nonspecific binding. Under such conditions and in a bi-molecular binding reaction, the concentration of the protein that leads to 50% substrate binding is the approximate dissociation constant. The nucleic acid probes used in EMSAs are typically radio or fluorescently labeled at their 5' or 3' ends to permit specific and sensitive detection. We prefer to use fluorescently labeled probes for DNA-based EMSA because fluorophore labeling of DNA substrates approaches 100% during the oligonucleotide synthesis step, fluorescent signals do not undergo decay as radio-labels do, and different fluorescent labels with non-overlapping emission spectra can be attached to each strand of the dsDNA to simultaneously monitor the binding and cleavage of both strands of DNA in the duplex. In addition to the nucleic acids, the protein parts of the ribonucleoprotein complex can also be fluorescent labeled via thio-chemistry to the cysteine side chains, thereby allowing the quantitative tracing of proteins in one single experiment. Due to the large molecular weight of the Cascade complex (~ 400 kDa), agarose was preferred over PAGE as the electrophoresis media for EMSAs in our lab. The more porous agarose gel allowed Cascade complexes to enter the gel fast, preventing them from precipitating in the loading well. The resulting Cascade/DNA bands were much sharper than the PAGE-separated bands.

Cascade-DNA interaction is usually transient and weak unless the DNA substrate contains an authentic target complementary to the crRNA spacer flanked by an interference PAM, which allows for stable R-loop formation. We performed EMSAs to evaluate the PAM preference of *Tfu*Cascade *in vitro*. A set of fluorophore-labeled 276-bp target DNAs containing various PAM sequences flanking the target site was PCR-amplified from the template plasmids. They were incubated at 2 nM concentration with an increasing concentration of *Tfu*Cascade, from 10 nM to 1 μ M. The complex was incubated at 37 °C for

30 mins, and the electrophoresis was performed at 4 °C to minimize complex dissociation. Surprisingly, we only observed a rather weak Cascade-dsDNA interaction, with an apparent dissociation constant at 200 nM. Considering that the preferred growth temperature of *Thermobifida fusca*'s is 55 °C rather than 37 °C, we explored the R-loop formation behavior of *Tfu*Cascade at different temperatures (Fig. 1A). When incubated at 58 °C for 30 mins, *Tfu*Cascade formed a tight complex with the cognate dsDNA target, with an apparent dissociation constant of ~15 nM (Fig. 1B). The complex formed at 58 °C migrated significantly faster than that at 37 °C, suggesting that they adopted two distinct conformations.

The following protocol describes the detailed procedure for the EMSAs.

1. dsDNA targets used in EMSA experiments were PCR amplified by using 5' fluorophore-labeled T7 primers. The 6-FAM labeled forward primer and the Cy5 labeled reverse primer annealed to 128-bp upstream and 113-bp downstream of the target site, respectively. pCDF-target plasmids used in the plasmid loss assay were used as PCR templates. Therefore, the resulting PCR products were Cy5 and 6FAM labeled at the 5' ends of the target strand and non-target strand DNA, respectively, and also contain 128 bp upstream of the PAM and 113 bp downstream of the protospacer.
2. Gel-purify the PCR products and determine their concentrations by measuring the absorbance at 260 nm.
3. Dilute the probe to 4 nM in a buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl, 5% glycerol. Mix 10 µl of the probe with 10 µl of Cascade at various concentrations, prepared in the same buffer.
4. Incubate the mixture at 58 °C for 30 mins.
5. Prepare a 1.5% agarose gel containing 1 x TAE (40 mM Tris, 20 mM acetic acid, 1 mM EDTA).
6. Load a 10 µl sample of each binding reaction directly onto the gel, without any loading dyes. Run the gel at 100 volts in 1 x TAE buffer at 4 °C for approximately 45 mins.
7. Scan the gel using a Typhoon 9200 scanner with the appropriate excitation and emission wavelength settings.

In addition to investigating target DNA binding by wild type or mutated Cascade, the EMSA can also be used to study complex formation between Cascade/R-loop and Cas3. The apparent dissociation constant determined using this method is ~60 nM (Fig. 1C). The following protocol describes the detailed procedure.

1. Incubate 100 nM of fluorophore-labeled probe with 120 nM of Cascade at 58 °C for 30 mins. This procedure enables assembly of fluorophore-labeled Cascade/R-loop with a concentration of 50 nM.
2. Dilute the fluorophore-labeled Cascade/R-loop to 4 nM using buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl, 5% glycerol.

3. Mix 10 μ l of the diluted fluorophore-labeled Cascade/R-loop with 10 μ l of Cas3 in the same buffer, Cas3 spans a concentration range from 0 nM to 640 nM.
4. Incubate the mixture at 4 °C for 60 mins.
5. Load a 10 μ l sample of each binding reaction onto a 1.5% agarose gel without any loading dyes. Run the gel using 1 x TAE buffer at 100 volts at 4 °C for approximately 90 mins.
6. Scan the gel using a Typhoon 9200 scanner with the appropriate excitation and emission wavelength settings.

Chemical probing

KMnO₄ footprinting is a widely used method to detect single-stranded regions within a duplexed DNA. It has been used to detect DNA unwinding during the initiation of DNA replication and the formation of the ssDNA transcription bubble inside the elongating RNA polymerase. The method is based on the selective oxidation of the nucleobase in thymidines in the single-stranded DNA. Strong bases such as piperidine can react with the modified nucleotides by ring opening and subsequent phosphodiester cleavage. In type I-E CRISPR-Cas systems, the R-loop forming process involves DNA unwinding, duplex formation between the crRNA spacer and target DNA strand. The displaced non-target DNA strand is detectable by KMnO₄ footprinting. The R-loop formation was thought to be initiated by PAM recognition and proceeds in a directional fashion, from the PAM-proximal to the -distal end. The sequence complementarity between crRNA and the initial eight nucleotides or so in the PAM-proximal region (seed-sequence region) of the DNA target has been shown to be more critical than the rest of the target region.

Our EMSA experiments revealed that *Tfu*Cascade/target DNA complex migrated to different locations when assembled at different temperatures. Room temperature incubation resulted in a fuzzy and slow-migrating band, whereas 58 °C incubation led to a sharp and fast-migrating band. Since the chemical composition of these two complexes were identical, we reasoned that their different migration behavior was due to conformational differences. We speculate room temperature incubation leads to an unstable dsDNA binding conformation, whereas 58 °C incubation results in full R-loop formation which bends the DNA into a more globular state. We performed KMnO₄ footprinting analysis to monitor the R-loop formation at different assembly conditions. In the NTS, there are two thymidine residues within the so called “seed sequence” region, and five more thymidines in the rest of the R-loop region. KMnO₄ footprinting cleaved all seven thymidines in the 58 °C incubated samples, suggesting that the entire R-loop is unwound in this condition (Fig. 1D). In contrast, no clear cleavage bands were observed when *Tfu*Cascade was incubated with the DNA substrate at 12 °C for one hour, suggesting that this incubation did not lead to any DNA unwinding (Fig. 1D). Serendipitously, we discovered that when the same sample was incubated at 12 °C for an extended amount of time (i.e. overnight), only the seed sequence region in dsDNA was unwound, as only the first two thymidine residues were cleaved (Fig. 1D). KMnO₄ footprinting therefore revealed that we could program the *Tfu*Cascade into three functional states, namely the PAM-searching state, the seed-bubble state, and the full R-loop state, by manipulating the incubation condition with the target dsDNA.

The following protocol describes the detailed procedure for the KMnO_4 footprinting experiment.

1. Incubate 3 nM of the 276-bp dsDNA substrate containing a 5'-6FAM label in the non-target strand DNA with 50 nM Cascade in 50 μl binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl, and 5% glycerol). Incubate samples at the specified temperature and duration. A PCR thermocycler can be used to control the incubation condition.
2. Add 5.5 μl of 16 mM KMnO_4 solution into the samples, mix gently and incubate 2 min at 30 $^{\circ}\text{C}$ to allow for oxidation of exposed thymidine residues in DNA.
3. Quench the permanganate modification by the addition of 6.5 μl of β -mercaptoethanol and 6.5 μl of 500 mM EDTA.
4. Extract samples with 70 μl phenol/chloroform by vigorous shaking on a vortex for 1 min. Centrifuge samples (10 min, 13,000 rpm) and take the aqueous layer on the top phase.
5. Precipitate samples with 2.5 volumes of absolute ethanol (-20°C). Centrifuge tubes (15 min, 13,000 rpm, 4°C) and carefully pipet away the supernatant.
6. Dissolve pellets with 50 μl of 10% (v/v) piperidine. Incubate samples at 90 $^{\circ}\text{C}$ for 30 min to induce cleavage at the modified residues.
7. Extract samples with 70 μl phenol/chloroform by vigorous shaking on a vortex for 1 min. Centrifuge samples (10 min, 13,000 rpm) and take the aqueous layer.
8. Precipitate samples with 2.5 volumes of absolute ethanol (-20°C). Centrifuge tubes (15 min, 13,000 rpm, 4°C) and carefully pipet away the supernatant. Take care not to disturb the precipitated pellet.
9. Air-dry samples for 30 minutes to get rid of ethanol residues, then dissolve the samples with 20 μl of electrophoresis loading buffer (50% formamide, 25 mM EDTA).
10. Separate samples on a 10% denaturing polyacrylamide gel, depending on the size of the DNA fragment. In our case, to map the cleavage site at single-nucleotide resolution, we generated a set of size markers using PCR-amplification of the same dsDNA substrates using the same 5'-6FAM labeled forward primer and a reverse primer starting from different positions along the substrate.
11. Fluorescent signals were recorded using a Typhoon 9200 scanner with the appropriate excitation and emission wavelength settings.

Assemble *Tfu*Cascade/seed-bubble, *Tfu*Cascade/R-loop, and *Tfu*Cascade/R-loop/Cas3 complexes for structural analysis

A thorough biochemical investigation enabled us to program the Type I-E ribonucleoprotein complexes into at least three functional states. The following experimental procedure was used to scale up the RNP preparation from nanomolar to sub-micromolar concentration

range, with special attention paid to ensure stoichiometric association of Cascade with dsDNA and Cas3 nuclease. The sequences of the oligos used in preparing DNA substrates are documented in Table S1. To prepare the oligonucleotide DNA duplex, target and non-target strands were dissolved into 100 μ M using buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl. The two strands were then mixed in a 1:1 molar ratio, and heated at 95 °C for 5 min and cool slowly to room temperature. *Tfu*Cascade/seed-bubble complex was assembled by mixing 1:1.3 molar ratio of *Tfu*Cascade and the annealed target DNA at room temperature for overnight. A slight excess of DNA substrate was critical to ensure that all *Tfu*Cascade molecules were occupied by dsDNA. As full R-loop formation induces a conformational change in *Tfu*Cascade, a mixture of *Tfu*Cascade molecules in the apo and R-loop states will hamper structure determination efforts. *Tfu*Cascade/R-loop complex was assembled by incubating *Tfu*Cascade/dsDNA complex at 58 °C for 30 min, followed by size exclusion chromatography to remove the excess DNA (Fig. 2A).

After R-loop formation, Cascade further recruits Cas3 for DNA nicking and degradation. Because Cas3 does not bind to the Cascade in the apo and seed-bubble states, it was critical to program most of the Cascade molecules into the full R-loop state. Even so, we observed that Cas3 was prone to dissociate from the Cascade/R-loop during biochemical reconstitution. Several extra steps were taken to ensure a stoichiometric amount of Cas3 is present in the assembled *Tfu*Cascade/R-loop/Cas3 complex (Fig. 2A). A Cas3 affinity pulldown was performed by mixing *Tfu*Cascade/R-loop with Twin-Strep-HRV-*Tfu*Cas3 in a 1:2 molar ratio at 4°C for 60 min, followed by Strep-tag purification to remove the unbound *Tfu*Cascade/R-loop. The Twin-Strep tag was then removed via PreScission protease cleavage, and size exclusion chromatography (SEC) was performed to remove the excess Cas3 (Fig. 2A). SEC, native-PAGE, SDS-PAGE, and urea-PAGE were performed to analyze the quality of the assembled *Tfu*Cascade/R-loop and *Tfu*Cascade/R-loop/Cas3 complexes. The elution volume of *Tfu*Cascade/R-loop complex (56.67 mL) was ~2.8 mL bigger than that of the *Tfu*Cascade/R-loop/Cas3 complex (53.91 mL), consistent with the formation of a bigger complex (Fig. 2B). The SEC-purified *Tfu*Cascade/R-loop and *Tfu*Cascade/R-loop/Cas3 complexes exhibited higher than 95% homogeneity on native PAGE, with the former migrating faster than the latter (Fig. 2C). In addition, SDS-PAGE analysis revealed stoichiometric amounts of Cas3 in the assembled *Tfu*Cascade/R-loop/Cas3 complex (1xCas3-1xCse1-2xCse2-1xCas5e-6xCas7-1xCas6e) (Fig. 2C). Urea-PAGE revealed the presence of stoichiometric amounts of crRNA and DNA target in both complexes, ~24% of the NTS was nicked in the assembled complexes (Fig. 2C).

Conclusion

The pipeline and assays outlined in this chapter lead to successful reconstitution of several complexes representing key events during the interference stage of type I-E CRISPR-Cas system. The complexes assembled using these methods exhibited high homogeneity and good stoichiometry, paving the way towards structure determination using Cryo-electron microscopy. Manipulating temperature enabled us to program the *Tfu*Cascade/dsDNA into different conformations. Using EMSA and chemical probing approaches, we were able to analyze DNA binding and R-loop formation by the *Tfu*Cascade. Full R-loop formation is necessary for Cas3 association, and we used affinity pulldown to enrich the *Tfu*Cascade/R-

loop/Cas3 complex. Extensive quality analysis, including SDS-PAGE, EMSA, and urea-PAGE, were performed across each step of assembly, to ensure the protein and nucleic acids components were assembled with correct stoichiometry and high homogeneity. This study provides invaluable lessons for characterizing target DNA binding and R-loop formation in other types of CRISPR-Cas systems. We also believe that our methods will serve as a reference for the studies of many other single or multi-component protein-nucleic acid complexes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work is supported by National Institutes of Health (NIH) grants GM 118174 and GM116632 to A.K. Y. X was supported by the Project Program of State Key Laboratory of Natural Medicines, China Pharmaceutical University.

Reference

- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, ... & Horvath, P (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709–1712. [PubMed: 17379808]
- Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, ... & Van Der Oost J (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891), 960–964. [PubMed: 18703739]
- Dillard KE, Brown MW, Johnson NV, Xiao Y, Dolan A, Hernandez E, ... & Ke, A (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell*.
- Hayes RP, Xiao Y, Ding F, Van Erp PB, Rajashankar K, Bailey S, ... & Ke, A (2016). Structural basis for promiscuous PAM recognition in type I–E Cascade from *E. coli*. *Nature*, 530(7591), 499. [PubMed: 26863189]
- Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, Nogales E, & Doudna JA (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences*, 111(18), 6618–6623.
- Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, ... & Zhang, R (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature structural & molecular biology*, 21(9), 771.
- Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, ... & Wiedenheft B (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*, 345(6203), 1473–1479. [PubMed: 25103409]
- Jore MM, Lundgren M, Van Duijn E, Bultema JB, Westra ER, Waghmare SP, ... & Beijer MR (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology*, 18(5), 529.
- Jung C, Hawkins JA, Jones SK, Xiao Y, Rybarski JR, Dillard KE, ... & Ke A (2017). Massively parallel biophysical analysis of CRISPR-Cas complexes on next generation sequencing chips. *Cell*, 170(1), 35–47. [PubMed: 28666121]
- Makarova KS, & Koonin EV (2015). Annotation and classification of CRISPR-Cas systems In *CRISPR* (pp. 47–75). Humana Press, New York, NY.
- Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, ... & Horvath P (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13(11), 722. [PubMed: 26411297]

- Mojica FJ, García-Martínez J, & Soria E (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), 174–182. [PubMed: 15791728]
- Mojica FJ, Díez-Villaseñor C, García-Martínez J, & Almendros C (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3), 733–740. [PubMed: 19246744]
- Mulepati S, Héroux A, & Bailey S (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, 1256996.
- Mulepati S, & Bailey S (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *Journal of Biological Chemistry*, jbc-M113.
- Pourcel C, Salvignol G, & Vergnaud G (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(3), 653–663. [PubMed: 15758212]
- Rutkauskas M, Sinkunas T, Songailiene I, Tikhomirova MS, Siksnys V, & Seidel R (2015). Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. *Cell reports*, 10(9), 1534–1543. [PubMed: 25753419]
- Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, ... & Zhang F (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*, 60(3), 385–397. [PubMed: 26593719]
- Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, & Siksnys V (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *The EMBO journal*, 32(3), 385–394. [PubMed: 23334296]
- Westra ER, van Erp PB, Künne T, Wong SP, Staals RH, Seegers CL, ... & de Vos WM (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell*, 46(5), 595–605. [PubMed: 22521689]
- Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, ... & Nogales E (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, 477(7365), 486. [PubMed: 21938068]
- Xiao Y, Luo M, Hayes RP, Kim J, Ng S, Ding F, ... & Ke A (2017). Structure basis for directional R-loop formation and substrate handover mechanisms in type I CRISPR-Cas system. *Cell*, 170(1), 48–60. [PubMed: 28666122]
- Xiao Y, Luo M, Dolan AE, Liao M, & Ke A (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science*, eaat0839. [PubMed: 29880725]
- Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, ... & Wang Y (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature*, 515(7525), 147. [PubMed: 25118175]

Fig. 1. Biochemical investigation of Type I-E CRISPR-Cas system from *Thermobifida fusca*. (A) EMSA analysis of temperature-dependent R-loop formation behavior by *Tfu*Cascade. The lower, middle, and upper bands represent the free dsDNA, Cascade/R-loop and Cascade/dsDNA, respectively. (B) EMSA analysis of *Tfu*Cascade and dsDNA target interaction at 58 °C, with an apparent binding constant of 15 nM. (C) EMSA analysis of *Tfu*Cascade/R-loop and Cas3 interaction, with an apparent binding constant of 60 nM. The lower and upper bands represent the Cascade/R-loop and Cascade/R-loop/Cas3 complexes, respectively. (D) KMnO₄ footprinting analysis of the differently incubated *Tfu*Cascade/dsDNA complexes. Bottom: schematics of different extents of DNA unwinding, 5'-NTS fluorescent label in red asterisk, and the probing sites in black arrows.

Fig. 2. Large scale *Tfu*Cascade/R-loop/Cas3 reconstitution and analysis. (A) Reconstitution and purification scheme of the *Tfu*Cascade/R-loop/Cas3 ternary complex. (B) Size exclusion chromatography (HiLoad 16/600 Superdex 200 pg) profiles of the *Tfu*Cascade/R-loop and *Tfu*Cascade/R-loop/Cas3 complexes. (C) Quality analysis of the ternary complex using native-polyacrylamide gel electrophoresis (PAGE), SDS-PAGE, and urea-PAGE.