ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Formation and Stoichiometry of CRISPR-Cascade Complexes with Varying Spacer Lengths Revealed by Native Mass Spectrometry

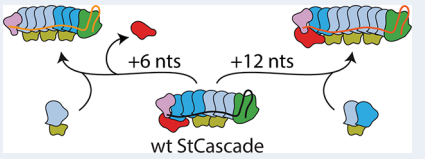
Sabine Wittig, Inga Songailiene, and Carla Schmidt*

Cite This: https://dx.doi.org/10.1021/jasms.9b00011

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The adaptive immune system of bacteria and archaea against viral DNA is based on clustered, regularly interspaced, short palindromic repeats (CRISPRs) which are encoded in the host genome and translated into CRISPR RNAs (crRNAs) containing single spacer sequences complementary to foreign DNA. crRNAs assemble with CRISPR-associated (Cas) proteins forming surveillance complexes that base-pair with viral DNA and mediate its degradation. As specificity of degradation is provided by the crRNA spacer sequence, genetic engineering of the CRISPR system has emerged as a popular molecular tool, for instance, in gene silencing and programmed DNA degradation. Elongating or shortening the crRNA spacer sequence are therefore promising ventures to modify specificity toward the target DNA. However, even though the stoichiometry of wild-type complexes is well established, it is unknown how variations in crRNA spacer length affect their stoichiometry. The CRISPR-associated antiviral defense surveillance complexes of *Streptococcus thermophilus* (StCascade complexes) contain crRNA and five protein subunits. Using native mass spectrometry, we studied the formation and stoichiometry of StCascade complexes assembled on a set of crRNAs with different spacer lengths. We assigned all relevant complexes and gained insights into the stoichiometry of the complexes as well as their preferred assembly. We found that stable complexes, which incorporate or lose a $(Cas7)_2(Cse2)_1$-module, assemble on crRNA varied in length by 12-nucleotide units, while varying crRNA length in six-nucleotide units results in heterogeneous mixtures of complexes. Combining our results from the various variants, we generated an assembly pathway revealing general features of I-E type Cascade complex formation.



KEYWORDS: *CRISPR, type I-E Cascade complex, crRNA, stoichiometry, complex heterogeneity, native mass spectrometry*

INTRODUCTION

Bacteria and archaea developed an adaptive immune system against viral DNA encoded by clustered, regularly interspaced, short palindromic repeats (CRISPRs) and CRISPR-associated proteins (Cas proteins).[1,2] After infection with foreign DNA, DNA spacer sequences are inserted into the CRISPR locus of the host genome.[3,4] The host transcription machinery then produces CRISPR RNAs (crRNAs) which contain single spacer sequences flanked by repeat sequences. crRNAs assemble with one or more Cas proteins to form surveillance complexes that base-pair with complementary foreign DNA followed by degradation of the invader, for instance, by Cas3 helicase-nuclease in Type I systems.[2,5,6] Selection of spacer sequences from invader DNA after initial infection is guided by the presence of protospacer adjacent motifs (PAMs) in the foreign genome.[7] The CRISPR locus of the host is lacking these sequences and is therefore protected against self-degradation.

Depending on the architecture of the CRISPR locus, CRISPR-Cas systems have been divided into two main classes and six types that are divided into many subtypes.[8] The most abundant system is the Type I system which is found in both bacteria and archaea[8,9] and recognizes double-stranded DNA. The Type I-E CRISPR-associated antiviral defense surveillance complexes (Cascade) of *Escherichia coli* and *Streptococcus Thermophilus* (*S. Thermophilus*) contain five Cas proteins in varying stoichiometry, namely, two copies of Cse2, six copies of Cas7, and one copy of the Cse1, Cas5, and Cas6 subunits.[10−12] Structurally, the Cas7 "backbone" (six copies) positions the crRNA,[13−15] and the two Cse2 copies stabilize crRNA and DNA base-pairing.[5,16,17] The large Cse1 subunit recognizes the PAM sequence of the DNA invader, and Cas5 and Cas6 cap the 5′ end and the 3′ stem loop of the crRNA, respectively.[10,12,17]

Typically, crRNAs comprise spacers of 32−38 base-pairs and repeats of 28−37 base-pairs.[18] Engineering crRNAs varying in spacer length is promising to modify specificity toward the target DNA. Accordingly, Cascade complexes with elongated or shortened crRNA spacers have been reported for *E. coli*,[19,20] *Shewanella putrefaciens*,[21] and *S. thermophilus*.[22] A recent study employing crRNA spacers up to +57 nucleotides showed that elongated spacer length increased DNA binding affinities and supports R-loop formation, although DNA degradation does

Received: September 26, 2019
Revised: January 18, 2020
Accepted: January 21, 2020

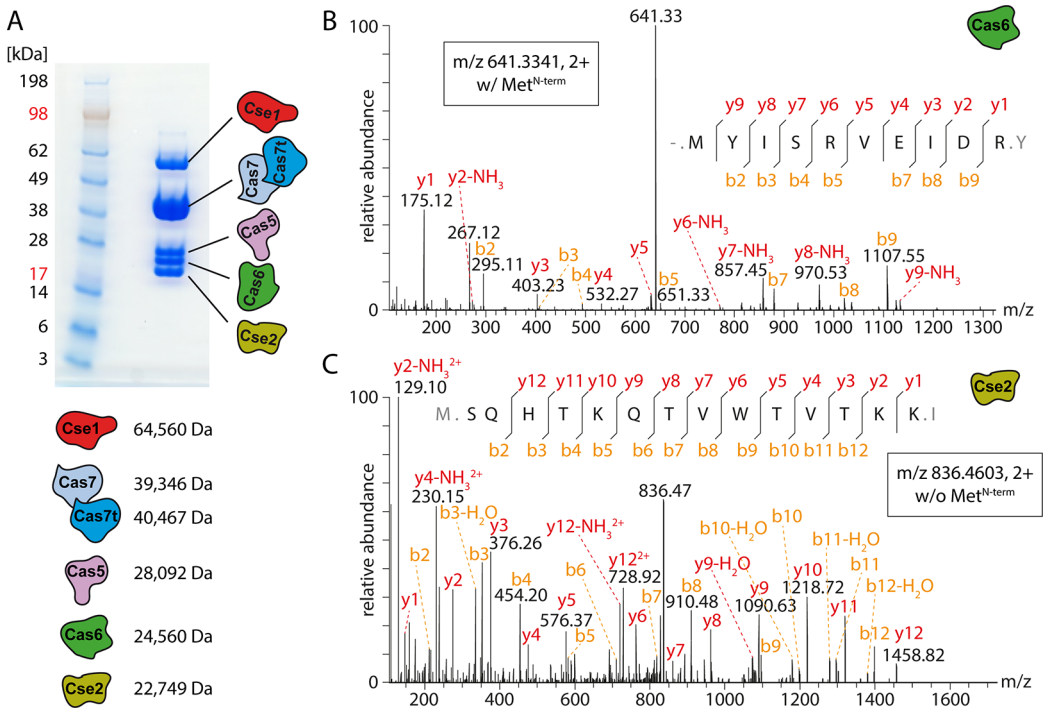


Figure 1. Identification of StCascade protein subunits. (A) SDS-PAGE of the purified wild-type complex revealed the presence of the five subunits. The accurate molecular weight was determined from LC-MS/MS experiments. (B) Fragment spectrum of the N-terminal peptide of Cas6 containing N-terminal methionine. y- and b-ion series are labeled (red and yellow). (C) Fragment spectrum of the N-terminal peptide of Cse2 lacking N-terminal methionine. y- and b-ion series are labeled (red and yellow).

not require extended R-loops and occurs for all R-loops corresponding to the wild-type.[22] Remaining uncertainties are, however, the stoichiometries and the stability of the assembled Cascade complexes. Previous studies revealed that one copy of Cas7 requires space of 6 nucleotides on the crRNA and Cse2 is equivalent to 12 nucleotides.[23] As Cas7 and Cse2 assemble on opposite sides of the crRNA, incorporation of one Cse2 and two Cas7 subunits per 12-nucleotide unit was therefore suggested.[19,20] Extension or shortening crRNA spacers by spacer units that are shorter or longer than 12 nucleotides, on the contrary, is less established. Previous studies suggested incorporation or loss of one Cas7 subunit per six nucleotide differences; however, these studies used Cascade complexes which omitted the large Cse1 subunit or lost the affinity-tagged subunit during purification suggesting dissociation of the complexes in solution.[19] It is therefore still unclear whether only Cas7 subunits are integrated in these complexes or if Cse2 subunits are also affected.

To gain insights into the formation and stoichiometries of Cascade complexes containing crRNAs differing in length, we studied a full set of StCascade complexes assembled in crRNAs containing spacers shortened or elongated in 6 nucleotide units ranging from −18 up to +24 nucleotides when compared with the wild-type crRNA. Using native mass spectrometry, we were able to detect the intact, assembled complexes and obtained insights into the presence of heterogeneous populations. Combining the results obtained from all StCascade complexes, we were able to unravel the assembly pathways of Cascade complexes on the various crRNAs and propose their preferred stoichiometries to form stable Cascade complexes.

METHODS

Purification of CRISPR-Cascade Complexes. Cascade complexes from *S. thermophilus* were purified as described.[11,22] Briefly, three compatible, heterologous plasmids containing the five protein subunits (Cse1, Cse2, Cas5, Cas6, Cas7), a C-terminally His-tagged Cas7 variant and the crRNA were coexpressed in *E. coli* BL21(DE3) cells, and the assembled StCascade complexes were purified through a His-tag strategy as described previously.[11] The wild-type complex assembled on a 61 nucleotides long crRNA comprising the 33 nucleotides long spacers as well as seven and 21 nucleotides long handles on the 5′ and 3′ ends, respectively. Elongated and shortened crRNA spacers were based on the wild-type sequences.[22] See Table S1 for masses of crRNAs.

LC-MS/MS and Database Searching. CRISPR-Cascade protein subunits were separated by gel electrophoresis using the NuPAGE system (Thermo Fisher Scientific) according to manufacturer's protocols. Protein bands were cut, and proteins were digested in gel as described.[24,25] The peptide mixture was separated by nanoflow reversed-phase liquid chromatography (DionexUltiMate 3000 RSLCnano System, Thermo Scientific) using 0.1% (v/v) formic acid as mobile phase A and 80% (v/v) acetonitrile/0.1% (v/v) formic acid as mobile phase B. The peptides were first loaded onto a trap column (μ-Precolumn C18 PepMap 100, C18, 300 μm I.D., particle size 5 μm; Thermo Scientific) and then separated at a flow rate of 300 nL/min and a gradient of 4−90% mobile phase B over 69 min using an analytical column (Acclaim PepMap 100, C18, 50 cm, 75 μm I.D., particle size 3 μm; Thermo Scientific).

Peptides were directly eluted into a Q Exactive plus hybrid mass spectrometer (Thermo Scientific) and analyzed by data-dependent acquisition. MS settings were: capillary voltage, 2.8 kV; capillary temperature, 275 °C; normalized collision energy,

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

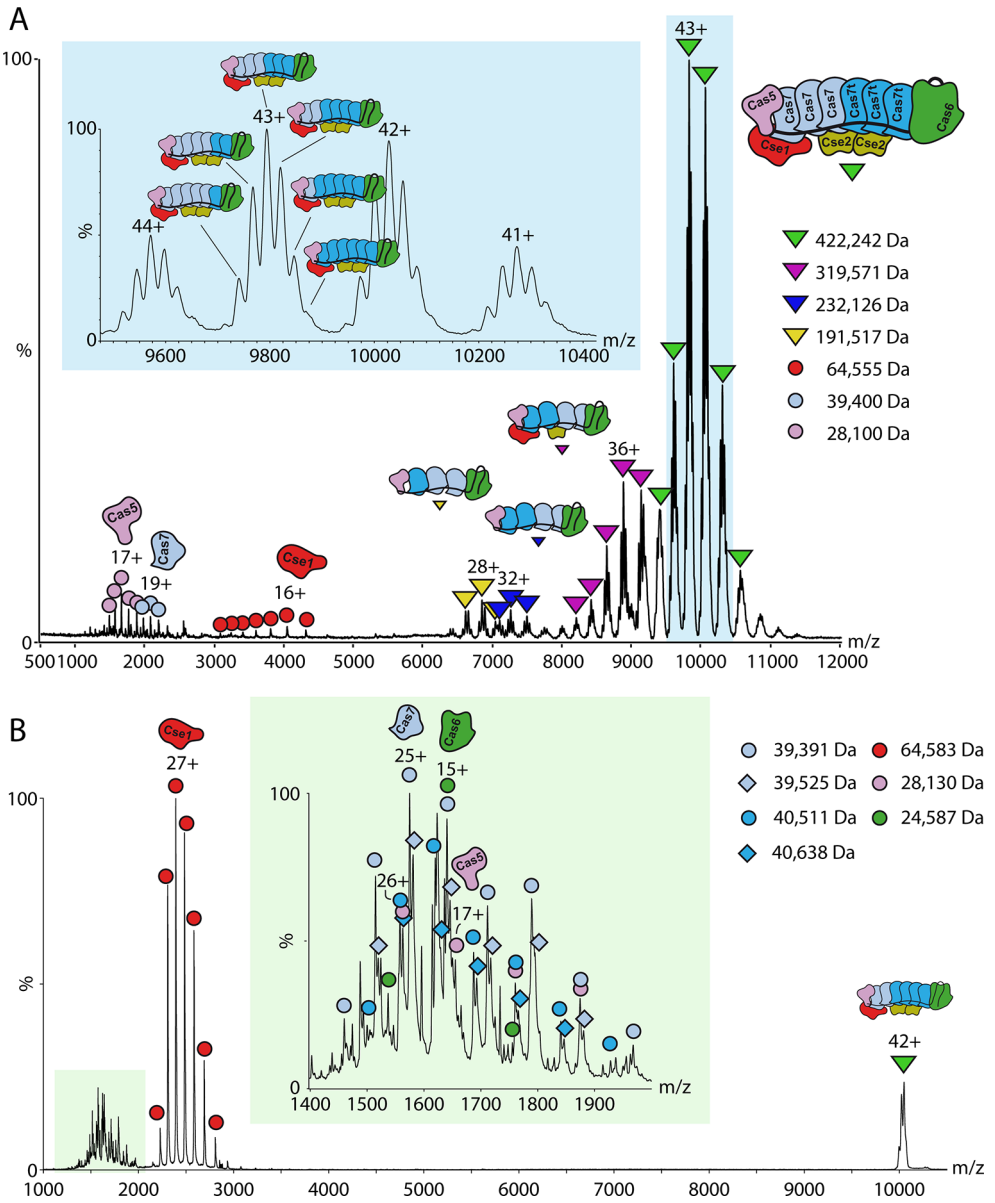


Figure 2. Native mass spectrometry of the wild-type StCascade complex. (A) The charge state distribution at approximately 10 000 m/z corresponds to the intact wild-type complex (green triangles). Subcomplexes (yellow, blue, and purple triangles) lost Cse1, Cse2, or Cas7 subunits and were observed between 6500 and 9500 m/z. Individual protein subunits were observed in the low m/z-region (<5000 m/z). See also Figure S1. (B) MS/MS spectrum of the 42+ charge state of the intact StCascade wild-type complex (m/z 10 049). Cse1 dissociates readily from the complex. Cas5, Cas6, and Cas7 subunits were observed at lower intensities. Collisional voltage: 180 V.

30%. Full MS scans (m/z 350−1600) were acquired in the Orbitrap at 70 000 resolution and with an automatic gain control target of 3e6. The 20 most intense ions were selected for fragmentation in the HCD cell with an automatic gain control target of 1e5. Selected precursors were dynamically excluded for 30 s before repeated fragmentation. Only ions with charge states of 2+ to 7+ were selected for fragmentation, singly charged ions, and ions with undetermined charge states were excluded. The Orbitrap was internally calibrated using the lock mass m/z 445.120025.[26]

Proteins were identified by database searching using the Mascot search engine v4.0.1. For this, raw data files were converted to mascot generic format. Search settings were: database, NCBInr (112 926 358 sequences); taxonomy, eubacteria (81 588 453 sequences); enzyme, trypsin; missed cleavages, up to 2; peptide charge, 2+, 3+, and 4+; precursor mass tolerance, 0.5 Da; mass tolerance of fragment ions, 10 ppm; variable modifications, carbamidomethylation of cysteines and oxidation of methionine.

Native MS and Data Analysis. Prior to analysis, the purification buffer of StCascade complexes was exchanged against 200 mM ammonium acetate pH 7.0 using Micro Bio-Spin 6 gel filtration columns (Bio-Rad). Intact protein−crRNA complexes were subsequently analyzed on an Ultima Q-ToF mass spectrometer (Micromass) modified for transmission of high mass complexes.[27] For this, 2−3 μL of protein sample was loaded in gold-coated emitter needles prepared in-house.[28] MS settings were: capillary voltage, 1.5−1.7 kV; cone voltage, 80 V; RF lens voltage, 80 V; collision voltage, 40−100 V. MS/MS settings were: capillary voltage, 1.5−1.7 kV; cone voltage, 80 V; RF lens voltage, 80 V; collision voltage, 40−200 V. Mass spectra were processed using MassLynx 4.0 and analyzed using

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

Massign software.[29] Mass spectra were calibrated externally using 100 mg/mL cesium iodide solution.

RESULTS

Characterization of the *S. thermophilus* CRISPR-Cas Wild-Type Complex. CRISPR-Cas complexes from *S. thermophilus* (StCascade complexes) were overexpressed in *E. coli* and purified through a His-tag strategy.[11,30] We first targeted the wild-type complex which is composed of crRNA containing a 33 nucleotide long spacer as well as five protein subunits. Using liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) and database searching after separation of the proteins by gel electrophoresis (Figure 1), we identified the five subunits (Cse1, Cse2, Cas5, Cas6, and Cas7) with high confidence as confirmed by high sequence coverage (>50% for Cse2 and >90% for all other subunits; note that Cse2 contains only a limited number of tryptic cleavage sites, and the sequence that can be covered after tryptic digestion was identified) and high peptide scores (max. peptide score >137) (Table S2).

To date, it is unknown whether StCascade protein subunits contain or omit N-terminal methionine. We therefore inspected the LC-MS/MS results for the existing protein sequence and identified StCascade protein subunits with or without N-terminal methionine. N-Terminal peptides identified for subunits Cas5 and Cas6 contained methionine, while subunits Cse1 and Cse2 did not contain methionine and Cas7 N-terminal peptides were identified with and without methionine (Figure 1B,C and Table S2).

We then analyzed the stoichiometry as well as homogeneity of the intact wild-type complex. For this, we used native mass spectrometry which maintains noncovalent interactions in the gas phase of the mass spectrometer.[31] The obtained mass spectrum shows several charge state distributions of intact StCascade complexes as well as individual proteins in the low mass region (<5000 m/z) (Figure 2A). In agreement with LC-MS/MS experiments, Cas7 (tagged and untagged, see Materials and Methods) showed two populations with small mass differences corresponding to the proteins with and without N-terminal methionine. However, peak intensities revealed that the main proportion of Cas7 (tagged and untagged) does not contain N-terminal methionine (Figure S1). Cas5 and Cas6 showed only one population containing N-terminal methionine (Figure S1). We therefore calculated all protein subunit masses according to these results (Figure 1) and used these masses for the assignment of identified Cascade complexes (Tables S3 and S4).

The major peak distribution in the native mass spectrum of the wild-type complex corresponds to a mass of 422 242 Da (Figure 2A) representing the intact StCascade complex with a stoichiometry of $(Cse1)_1(Cse2)_2(Cas7)_6(Cas5)_1$-$(Cas6)_1(crRNA)_1$ as described previously.[10,12,32] Close inspection of the individual charge states of this peak distribution reveals a distribution of complexes differing in mass by approximately 1100 Da corresponding to the presence of multiple His-tagged Cas7 subunits (Figure 2A, inset). In agreement with the identified stoichiometry, charge state series of the wild-type complex contained one to up to six tagged Cas7 subunits confirming the calculated protein stoichiometry (Table S3).

To study stability and complex assembly of the intact wild-type complex, we selected the 42+-charge state for tandem MS experiments (MS/MS) and found that the Cse1 subunit readily dissociates from the complex confirming its peripheral position in the complex (Figure 2B). Dissociated, highly charged Cas7, Cas5, and Cas6 subunits were also identified, albeit at lower intensities. We further studied complex stability at varying collisional energies and found that the wild-type complex is rather stable at high collisional energies; The complexes dissociated only at elevated collisional voltages (Figure S2). Sequential dissociation of the subunits was not observed suggesting that Cse1 stabilizes the complexes and, after dissociation of Cse1, overall binding of other subunits is weakened.

Besides the major peak distribution of the intact wild-type complex, we observed additional charge state series of smaller subcomplexes. These were generated by loss of peripheral subunits Cse1 and Cse2 as well as tagged and untagged Cas7 subunits. Notably, the second most intense peak distribution corresponds to a subcomplex which lost two copies of Cas7 and one copy of Cse2 that appears to be a stable module of CRISPR-Cascade complexes (see below). Other subcomplexes that lost additional copies of Cas7 as well as Cse1 and Cse2 subunits were observed at lower intensities. Notably, both Cas5 and Cas6 subunits did not dissociate from the wild-type complex suggesting that they are stably attached to the 5′ and 3′ end of the crRNA.

StCascade Complexes Assembled on crRNA Differing in Length by ±12 Nucleotides Are Stable and Show Defined Stoichiometries. To explore whether longer or shorter crRNAs allow the assembly of stable StCascade complexes, we elongated or shortened the crRNA spacer length in 6- nucleotide units (−18, −12, −6, +6, +12, +18, +24 nucleotides). We first targeted StCascade complexes assembled on crRNAs that differ in 12-nucelotide units when compared with the wild-type complex, i.e., crRNA spacer length of +12, +24, and −12 nucleotides. For these complexes incorporation or loss of two Cas7 and one Cse2 copies per 12 nucleotides was suggested.[20]

StCascade complexes were purified and analyzed as described (Methods). The native mass spectrum of the StCascade complex assembled on 12 nucleotides shorter crRNA showed three species: one major species, which comprises the expected stoichiometry of $(Cse1)_1(Cse2)_1$-$(Cas7)_4(Cas5)_1(Cas6)_1(crRNA)_1$ and two subcomplexes thereof, which lost the peripheral subunit Cse1 as well as Cse1 and Cse2 subunits and showed stoichiometries of $(Cse2)_1(Cas7)_4(Cas5)_1(Cas6)_1(crRNA)_1$ and $(Cas7)_4(Cas5)_1(Cas6)_1(crRNA)_1$ (Figure 3A). Loss of Cse1 and Cse2 subunits was also observed for subcomplexes of the wild-type StCascade complex. Their functional role of binding invader DNA during CRISPR-Cas mediated DNA degradation suggests tight binding in functionally active complexes and a rather loose association in the absence of double-stranded DNA as indicated by the presence of subcomplex which lost these subunits.

We confirmed the assignment of these complexes by MS/MS of the full complex as well as a subcomplex. Similar to the wild-type complex, we found that Cse1 readily dissociates. Highly charged Cas7 monomers were also identified in the low m/z region of the mass spectrum. The subcomplex which does not contain Cse1, on the other hand, only showed dissociation of Cas7 and Cas5 subunits.

In addition to shorter crRNAs, longer crRNAs were also explored. When StCascade complexes were assembled on crRNA which is 12 nucleotides longer than wild-type crRNA,

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

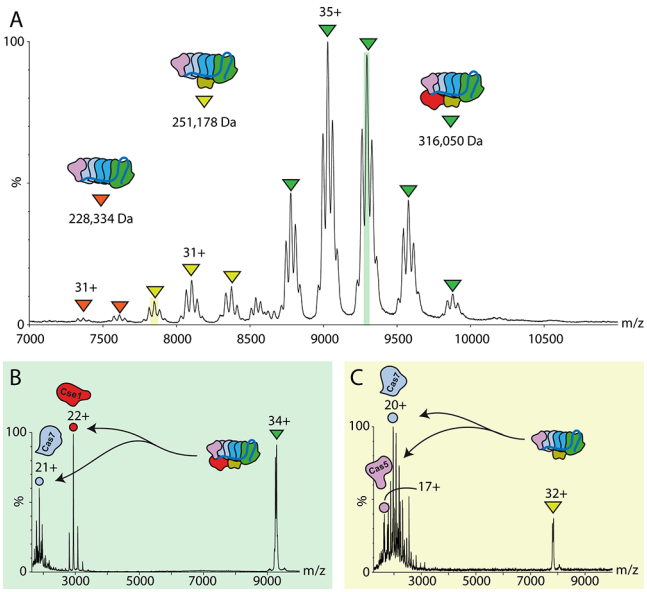


Figure 3. Native mass spectrometry of the StCascade complex assembled on crRNA containing a 12 nucleotides shorter spacer. (A) A complex with the expected stoichiometry of $(Cse1)_1(Cse2)_2$-$(Cas7)_4(Cas5)_1(Cas6)_1(crRNA)_1$ was observed. Subcomplexes that lost Cse1 or Cse2 subunits were also observed. (B) MS/MS spectrum of the major complex. The 34+ charge state (m/z 9285) was selected for fragmentation. Highly charged Cse1 and Cas7 monomers were observed in the low m/z region of the spectrum. Collisional voltage: 150 V. (C) MS/MS spectrum of the $(Cse2)_1(Cas7)_4(Cas5)_1(Cas6)_1$-$(crRNA)_1$ subcomplex. The 32+ charge state (m/z 7844) was selected for fragmentation. Monomeric Cas7 and Cas5 were found to dissociate from the complex. Collisional voltage: 150 V.

one major peak distribution between 11 000 and 13 000 m/z was observed (Figure S3). In accordance with previous studies, one Cse2 and two Cas7 subunits were incorporated into this complex yielding a stoichiometry of $(Cse1)_1(Cse2)_3$-$(Cas7)_8(Cas5)_1(Cas6)_1(crRNA)_1$. Again, the cohort of Cas7 subunits comprises a mixture of tagged and untagged subunits. The resolution obtained for this half-megadalton complex is sufficient to assign the multiple populations with high accuracy (Figure S3 and Table S3).

In a similar fashion, a high-mass complex of 636 000 Da was observed for the StCascade complex assembled on 24 nucleotides longer crRNA. In agreement with addition of the $(Cas7)_2(Cse1)$ module (see above) and an increase in crRNA mass, a stoichiometry of $(Cse1)_1(Cse2)_4(Cas7)_{10}(Cas5)_1$-$(Cas6)_1(crRNA)_1$ was calculated for this complex (Table S3). Because of the number of protein subunits, the resulting high mass of the intact complex, the presence of up to 10 tagged Cas7 subunits and the presence of many coexisting overlapping species, the observed mass spectrum is not sufficiently resolved, and it is difficult to assign all subcomplexes and populations. Nonetheless, several possible subcomplexes of the fully assembled complex could be assigned. The high intensity of the intact complex indicates its preferred formation (Figure S4).

In summary, the mass spectra of StCascade complexes assembled on crRNAs with decreased or increased spacer length (−12, +12, and +24 nucleotides) confirm the incorporation or loss of the $(Cas7)_2(Cse1)$ module for each 12-nucleotide unit, as suggested previously.[20] The increasing number of protein subunits including tagged and untagged populations of Cas7 as well as overlapping species prove the assignment of all subcomplexes assembled on longer crRNAs difficult. Nonetheless, we were able to assign the major complex species with high confidence. Importantly, the distribution of populations containing tagged/untagged Cas7 subunits helped assign the full complexes as well as possible subcomplexes. The comparably high intensity of the fully assembled complexes indicates their stability in solution and confirms the preferred stoichiometries of fully assembled StCascade complexes.

Increasing or Decreasing crRNA Spacer Length by Odd-Numbered Six Nucleotide Units Causes Formation of Heterogeneous StCascade Complex Mixtures. We next investigated StCascade complexes assembled on crRNAs extended or shortened by odd-numbered 6-nucleotide units when compared with their 12-nucleotides counterparts (−18, −6, +6, and +18 nucleotides). Incorporation and loss of one Cas7 copy, respectively, was suggested for these complexes;[19,20] however, it is unclear whether the copy number of Cse2 subunits is affected. We started analyzing the StCascade complex assembled on crRNA which was six nucleotides shorter than the crRNA of the wild-type complex. The observed mass spectrum revealed a complex with the stoichiometry of $(Cse2)_2(Cas7)_5(Cas5)_1(Cas6)_1(crRNA)_1$. This complex agrees well with the suggested loss of one Cas7 subunit; however, it also revealed additional loss of Cse1. The absence of Cse1 was confirmed by MS/MS (Figure 4);

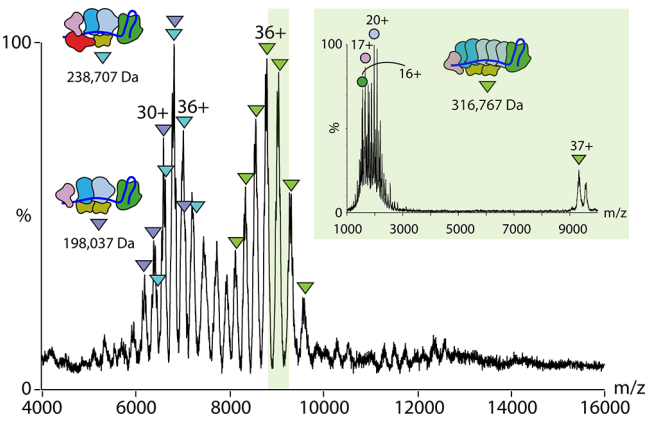


Figure 4. Native mass spectrometry of the StCascade complex assembled on crRNA containing a six nucleotide shorter spacer. A complex with the stoichiometry of $(Cse2)_2(Cas7)_5(Cas5)_1(Cas6)_1$-$(crRNA)_1$ was observed. Subcomplexes that lost Cse1, or Cse2 as well as Cas7 were also observed. Inset: MS/MS spectrum of the major complex. The 37+ charge state (m/z 9324) was selected for fragmentation. Cas5, Cas6, and Cas7 monomers dissociated from the complex. Collisional voltage: 180 V.

the respective spectrum showed the loss of Cas7, Cas5, and Cas6 subunits without dissociation of Cse1 which is preferred during MS/MS of complexes that contained Cse1 (see above). Additional subcomplexes, which lost Cas7 subunits or maintained Cse1 but lost one copy of Cse2, were also observed (Figure 4 and Table S4).

We also analyzed a complex that contained an 18 nucleotides shorter crRNA spacer when compared with wild-type complex. The native mass spectrum of this complex showed several overlapping peak distributions at sufficient resolution allowing assignment of all subcomplexes (Figure S5). Again, the major peak distribution corresponds to the

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

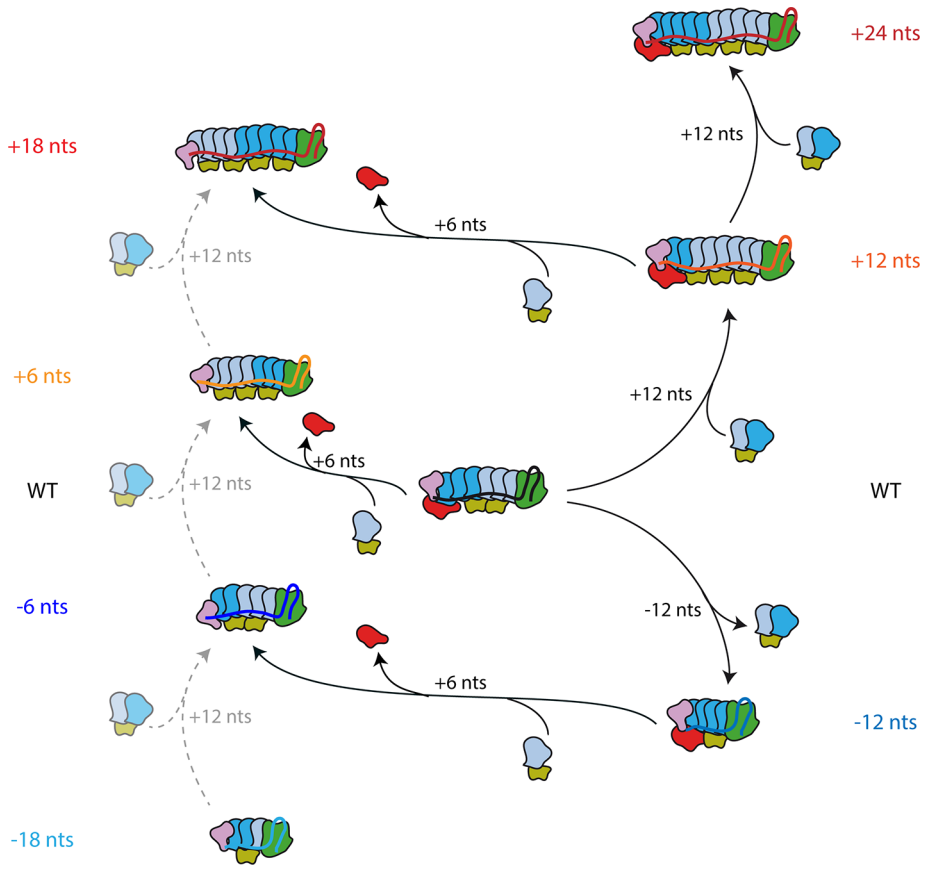


Figure 5. Assembly of StCascade complexes on crRNAs with altered spacer length. The wild-type complex (WT) is shown in the center. Increasing or decreasing the length of the crRNA spacer by 12 nucleotides causes incorporation or loss of the $(Cse2)_1(Cas7)_2$ module (rhs). Increasing the crRNA spacer length by six nucleotides yields heterogeneous populations of StCascade complexes presumably due to space constraints. The preferred assembly includes incorporation of one copy Cse2 and one copy Cas7 accompanied by the loss of Cse1 (lhs).

$(Cse2)_1(Cas7)_3(Cas5)_1(Cas6)_1(crRNA)_1$ complex which lost the Cse1 subunit in solution. In addition, we identified two subcomplexes, namely, $(Cas7)_3(Cas5)_1(Cas6)_1(crRNA)_1$ and $(Cse2)_1(Cas7)_3(Cas6)_1(crRNA)_1$, which lost Cas5 or Cse2 subunits (Table S4). Several populations containing tagged and untagged Cas7 subunits were observed for all complexes, and the number of tags confirmed the assignment of these complexes.

We then analyzed StCascade complexes assembled on longer crRNAs. The mass spectrum of the complex assembled on crRNA that contained a six nucleotide longer spacer (+6 nucleotides) showed overlapping peak distributions between 9000 and 12000 m/z (Figure S6). Two complexes with masses of 421 451 Da and 398 486 Da could be assigned, both present as populations with tagged and untagged Cas7 subunits. Complex stoichiometries of $(Cse2)_3(Cas7)_7(Cas5)_1(Cas6)_1$-$(crRNA)_1$ and $(Cse1)_1(Cse2)_2(Cas7)_6(Cas6)_1(crRNA)_1$ were calculated for these complexes (Table S4). A complex with the expected stoichiometry of $(Cse1)_1(Cse2)_2(Cas7)_7(Cas5)_1$-$(Cas6)_1(crRNA)_1$ corresponding in mass to approximately 462 kDa was not observed. Instead, the observed complexes were obtained by incorporation of one Cas7 and one Cse2 subunits, accompanied by loss of Cse1, or by dissociation of Cas5 presumably together with other subunits after assembly of larger complexes. The presence of multiple complexes that do not contain all protein subunits suggests that StCascade complexes assembled on crRNA extended by six nucleotides are not preferred and form rather loose assemblies.

The mass spectrum of StCascade containing 18 nucleotides longer crRNA revealed a rather unresolved peak distribution (Figure S7). Most likely this is caused by overlapping peak distributions of complexes with similar masses (see also above). However, the average mass obtained from this peak distribution corresponds to a complex with a stoichiometry of $(Cse2)_4(Cas7)_9(Cas5)_1(Cas6)_1(crRNA)_1$, which is equivalent to the complex described above for StCascade assembled on crRNA extended by six nucleotides (Table S4). More precisely, it represents incorporation of one Cas7 and one Cse2 subunit and simultaneous loss of Cse1 when compared with the counterpart complex assembled on +12 nucleotides longer crRNA. These findings suggest a general feature: Namely, incorporation of one Cas7 and one Cse2 subunits when the crRNA is extended by six nucleotides. Because of space constraints on the crRNA, the complex dissociates into subcomplexes of higher stability; loss of Cse1 is one of the preferred dissociation pathways most likely because the peripheral Cas5 and Cas6 subunits stably interact with the 5′ and 3′ end of the crRNA.

DISCUSSION

The structural assembly of CRISPR effector complexes has been investigated for several CRISPR subtypes in numerous bacterial and archaeal species. Most of the insights have been obtained from high resolution X-ray structures or electron microscopy images.[33] In many cases, mass spectrometry assisted structural analyses, for instance, by protein−protein

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

and protein−RNA cross-linking approaches (e.g., refs 21, 34, and 35) or by identification and quantification of Cas effector proteins (e.g., refs 34, 36, and 37). Native mass spectrometry is often applied when protein stoichiometries of effector complexes are investigated (e.g., refs 19, 20, 32, and 38). This is of particular interest as effector complexes containing crRNAs differing in length have been purified,[39,40] and changes in protein stoichiometries were suggested. In general, CRISPR-Cas has become a popular tool in molecular biology, and engineering of crRNA spacers, including variations in spacer length, is promising for modifying DNA specificity. The analysis of engineered effector complexes is therefore of great interest.

Here we used native mass spectrometry to investigate protein stoichiometries of StCascade complexes containing crRNA spacers varying in spacer length. Nuclease-mediated ribonucleic acid cleavage was previously described to occur in six nucleotides units.[39,40] We therefore employed a full-set of crRNAs containing spacers varying in length by −18, −12, −6, +6, +12, +18, and +24 nucleotides when compared with wild-type crRNA. Similar to previous studies, we found that the $(Cas7)_2(Cse2)_1$ module is preferably incorporated or lost per 12 nucleotides increment or decrement. The fully assembled complexes are observed with superior intensity, and we assume that they represent the most stable StCascade complexes. When increasing or decreasing crRNA length by less or more than 12 nucleotide units, complex mixtures are observed presumably due to structural mismatch; for instance, when elongating/shortening crRNA by six nucleotides, as performed in this study, a stable Cse2 backbone cannot be formed causing dissociation of other protein subunits. As a consequence, we found that Cse1 preferably dissociates in solution from crRNAs that do not accommodate the full $(Cas7)_2(Cse2)_1$ module (i.e., crRNAs that varied in length by −18, −6, +6, +18 nucleotides). In agreement with this, we also found that Cse1 readily dissociates from fully assembled complexes during MS/MS experiments confirming its peripheral position in the complex and suggesting loose association with the complex. Similar findings were described for the protein subunit Cmr1 of the type III-A Cmr effector complexes from *Thermus thermophilus*.[39]

To date, engineered crRNA spacers have been employed for type I-E cascade complexes of *E. coli*[19,20] and *S. thermophilus*[6] as well as type I-F cascade complexes from *S. putrefaciens*.[21] In all cases, the number of protein subunits that assembled on crRNAs with elongated or shortened spacers adjusted, resulting in Cascade complexes with modified stoichiometries. Reduced activity was, however, reported for most effector complexes assembled on shorter crRNA spacers.[19,21] For type I-E StCascade complexes, DNA cleavage activity was only observed for crRNA spacers corresponding in length to the wild-type or +12 and +24 nucleotides.[6] Our results support these findings as fully assembled complexes that contained all protein subunits were only observed for these StCascade complexes. For complexes assembled on +6 and +18 nucleotides longer crRNA spacers, heterogeneous mixtures of StCascade complexes were observed. As these Cascade complexes were found to show reduced activity,[6] we conclude that activity correlates with the presence of all protein subunits. A unique feature observed for −18, −6, +6, and +18 nucleotides varied crRNA spacers is dissociation of Cse1 from StCascade complexes. As Cse1 is important for PAM recognition, the absence of this protein subunit most likely causes loss of activity. The absence of Cse1 was also discussed in a previous study that showed that R-loop formation does not occur in the absence of Cse1.[19] We conclude that wild-type crRNA spacers of defined length are optimized for cascade assembly, and engineered spacers have to be optimized accordingly.

CONCLUSIONS

We show that native mass spectrometry is well-suited to study large and heterogeneous protein assemblies such as those assembled on crRNAs of varying lengths. Assignment of the spectra obtained from heterogeneous complex mixtures can be complex. Here, the presence of peak populations differing in the numbers of tagged Cas7 subunits helped assign the full complexes by confirming the total number of Cas7 subunits. When compared with other structural techniques, for instance, cryo-electron microscopy or X-ray crystallography, native mass spectrometry is straightforward, less laborious, and allows simultaneous analysis of heterogeneous populations of protein−ligand complexes. In summary, results obtained from native mass spectrometry together with biochemical assays provide insights into the activity of protein assemblies and help in the understanding of their modes of action.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.9b00011.

Supporting tables: Table S1: Masses of StCascade crRNAs, Table S2: Protein identification of StCascade protein subunits, Table S3: Stoichiometries of StCascade complexes assembled on crRNAs differing in spacer length by ±12 nucleotides, Table S4: Stoichiometries of StCascade complexes assembled on crRNAs differing in spacer length by ±6 nucleotides. Supporting figures: Figure S1: StCascade protein subunits, Figure S2: Mass spectra of the StCascade wild-type complex at varying collisional voltages, Figure S3: Native mass spectra of the StCascade complex assembled on crRNA containing a 12 nucleotides longer spacer than crRNA of the wild-type complex, Figure S4: StCascade complexes assembled on crRNA containing a 24 nucleotides longer spacers, Figure S5: StCascade complexes assembled on crRNA containing an 18 nucleotides shorter spacer than the wild-type complex, Figure S6: StCascade complex assembled on crRNA containing a 6 nucleotides longer spacer than crRNA of the wild-type complex, Figure S7: StCascade complexes assembled on crRNA containing a spacer which differs in length by +18 nucleotides (PDF)

AUTHOR INFORMATION

Corresponding Author

Carla Schmidt − *Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Institute for Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, 06120 Halle, Germany;* orcid.org/0000-0001-9410-1424; Email: carla.schmidt@biochemtech.uni-halle.de

Authors

Sabine Wittig − *Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Institute for Biochemistry and*

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX

Biotechnology, Martin Luther University Halle-Wittenberg, 06120 Halle, Germany

Inga Songailiene − *Department of Protein−DNA Interactions, Institute of Biotechnology, Vilnius University, 10257 Vilnius, Lithuania*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.9b00011

Author Contributions

S.W. performed all MS experiments and analyzed the data. I.S. purified and characterized Cascade complexes. C.S. analyzed the data and wrote the manuscript. All authors approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Professor Ralf Seidel (Leipzig University) and Professor Virginijus Siksnys (Vilnius University) for helpful discussions and support, Marius Rutkauskas (Leipzig University) for his contributions to figure cartoons, Dr. Tomas Sinkunas for providing plasmids and helpful discussions on purification of StCascade complexes, and Professor Nina Morgner (Frankfurt University) for helpful advice with data analysis. C.S. and S.W. acknowledge funding from the Federal Ministry for Education and Research (BMBF, ZIK programme, 03Z22HN22), the European Regional Development Funds (EFRE, ZS/2016/04/78115), and the MLU Halle-Wittenberg. I.S. acknowledges funding through an intramural Vilnius University grant and funding of a research internship to Leipzig University from the Research Council of Lithuania.

■ REFERENCES

(1) van der Oost, J.; Westra, E. R.; Jackson, R. N.; Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2014**, *12* (7), 479−92.

(2) Westra, E. R.; Swarts, D. C.; Staals, R. H.; Jore, M. M.; Brouns, S. J.; van der Oost, J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **2012**, *46*, 311−39.

(3) Amitai, G.; Sorek, R. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* **2016**, *14* (2), 67−76.

(4) Heler, R.; Marraffini, L. A.; Bikard, D. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* **2014**, *93* (1), 1−9.

(5) Hochstrasser, M. L.; Taylor, D. W.; Bhat, P.; Guegler, C. K.; Sternberg, S. H.; Nogales, E.; Doudna, J. A. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (18), 6618−23.

(6) Rutkauskas, M.; Sinkunas, T.; Songailiene, I.; Tikhomirova, M. S.; Siksnys, V.; Seidel, R. Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep.* **2015**, *10* (9), 1534−1543.

(7) Mojica, F. J.; Diez-Villasenor, C.; Garcia-Martinez, J.; Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **2009**, *155* (Pt 3), 733−40.

(8) Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J.; Backofen, R.; Koonin, E. V. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2015**, *13* (11), 722−36.

(9) Burstein, D.; Sun, C. L.; Brown, C. T.; Sharon, I.; Anantharaman, K.; Probst, A. J.; Thomas, B. C.; Banfield, J. F. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* **2016**, *7*, 10613.

(10) Jore, M. M.; Lundgren, M.; van Duijn, E.; Bultema, J. B.; Westra, E. R.; Waghmare, S. P.; Wiedenheft, B.; Pul, U.; Wurm, R.; Wagner, R.; Beijer, M. R.; Barendregt, A.; Zhou, K.; Snijders, A. P.; Dickman, M. J.; Doudna, J. A.; Boekema, E. J.; Heck, A. J.; van der Oost, J.; Brouns, S. J. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **2011**, *18* (5), 529−36.

(11) Sinkunas, T.; Gasiunas, G.; Waghmare, S. P.; Dickman, M. J.; Barrangou, R.; Horvath, P.; Siksnys, V. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *EMBO J.* **2013**, *32* (3), 385−94.

(12) Wiedenheft, B.; Lander, G. C.; Zhou, K.; Jore, M. M.; Brouns, S. J. J.; van der Oost, J.; Doudna, J. A.; Nogales, E. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **2011**, *477* (7365), 486−489.

(13) Jackson, R. N.; Golden, S. M.; van Erp, P. B.; Carter, J.; Westra, E. R.; Brouns, S. J.; van der Oost, J.; Terwilliger, T. C.; Read, R. J.; Wiedenheft, B. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science* **2014**, *345* (6203), 1473−9.

(14) Mulepati, S.; Heroux, A.; Bailey, S. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **2014**, *345* (6203), 1479−84.

(15) Zhao, H.; Sheng, G.; Wang, J.; Wang, M.; Bunkoczi, G.; Gong, W.; Wei, Z.; Wang, Y. Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature* **2014**, *515* (7525), 147−50.

(16) Sashital, D. G.; Wiedenheft, B.; Doudna, J. A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **2012**, *46* (5), 606−15.

(17) Hayes, R. P.; Xiao, Y.; Ding, F.; van Erp, P. B.; Rajashankar, K.; Bailey, S.; Wiedenheft, B.; Ke, A. Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature* **2016**, *530* (7591), 499−503.

(18) Barrangou, R.; Marraffini, L. A. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell* **2014**, *54* (2), 234−44.

(19) Kuznedelov, K.; Mekler, V.; Lemak, S.; Tokmina-Lukaszewska, M.; Datsenko, K. A.; Jain, I.; Savitskaya, E.; Mallon, J.; Shmakov, S.; Bothner, B.; Bailey, S.; Yakunin, A. F.; Severinov, K.; Semenova, E. Altered stoichiometry Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. *Nucleic Acids Res.* **2016**, *44* (22), 10849−10861.

(20) Luo, M. L.; Jackson, R. N.; Denny, S. R.; Tokmina-Lukaszewska, M.; Maksimchuk, K. R.; Lin, W.; Bothner, B.; Wiedenheft, B.; Beisel, C. L. The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. *Nucleic Acids Res.* **2016**, *44* (15), 7385−7394.

(21) Gleditzsch, D.; Muller-Esparza, H.; Pausch, P.; Sharma, K.; Dwarakanath, S.; Urlaub, H.; Bange, G.; Randau, L. Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res.* **2016**, *44* (12), 5872−82.

(22) Songailiene, I.; Rutkauskas, M.; Sinkunas, T.; Manakova, E.; Wittig, S.; Schmidt, C.; Siksnys, V.; Seidel, R. Decision-Making in Cascade Complexes Harboring crRNAs of Altered Length. *Cell Rep.* **2019**, *28* (12), 3157−3166.

(23) van Erp, P. B.; Jackson, R. N.; Carter, J.; Golden, S. M.; Bailey, S.; Wiedenheft, B. Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic Acids Res.* **2015**, *43* (17), 8381−91.

(24) Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856−2860.

(25) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole sequencing of

proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379* (6564), 466−9.
(26) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010−21.
(27) Sobott, F.; Hernandez, H.; McCammon, M. G.; Tito, M. A.; Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **2002**, *74* (6), 1402−7.
(28) Hernandez, H.; Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2007**, *2* (3), 715−26.
(29) Morgner, N.; Robinson, C. V. Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.* **2012**, *84* (6), 2939−48.
(30) Songailiene, I.; Rutkauskas, M.; Sinkunas, T.; Manakova, E.; Wittig, S.; Schmidt, C.; Siksnys, V.; Seidel, R. Decision making in Cascade complexes harboring crRNAs of altered length. *Cell Rep.* **2019**, *28*, 3157.
(31) Mehmood, S.; Allison, T. M.; Robinson, C. V. Mass spectrometry of protein complexes: from origins to applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 453−74.
(32) van Duijn, E.; Barbu, I. M.; Barendregt, A.; Jore, M. M.; Wiedenheft, B.; Lundgren, M.; Westra, E. R.; Brouns, S. J.; Doudna, J. A.; van der Oost, J.; Heck, A. J. Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa. *Mol. Cell. Proteomics* **2012**, *11* (11), 1430−41.
(33) Jiang, F.; Doudna, J. A. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **2015**, *30*, 100−111.
(34) Fagerlund, R. D.; Wilkinson, M. E.; Klykov, O.; Barendregt, A.; Pearce, F. G.; Kieper, S. N.; Maxwell, H. W. R.; Capolupo, A.; Heck, A. J. R.; Krause, K. L.; Bostina, M.; Scheltema, R. A.; Staals, R. H. J.; Fineran, P. C. Spacer capture and integration by a type I-F Cas1-Cas2−3 CRISPR adaptation complex. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (26), E5122−E5128.
(35) Hrle, A.; Maier, L. K.; Sharma, K.; Ebert, J.; Basquin, C.; Urlaub, H.; Marchfelder, A.; Conti, E. Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. *RNA Biol.* **2014**, *11* (8), 1072−82.
(36) Brendel, J.; Stoll, B.; Lange, S. J.; Sharma, K.; Lenz, C.; Stachler, A. E.; Maier, L. K.; Richter, H.; Nickel, L.; Schmitz, R. A.; Randau, L.; Allers, T.; Urlaub, H.; Backofen, R.; Marchfelder, A. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. *J. Biol. Chem.* **2014**, *289* (10), 7164−77.
(37) Plagens, A.; Tripp, V.; Daume, M.; Sharma, K.; Klingl, A.; Hrle, A.; Conti, E.; Urlaub, H.; Randau, L. In vitro assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic Acids Res.* **2014**, *42* (8), 5125−38.
(38) Rouillon, C.; Zhou, M.; Zhang, J.; Politis, A.; Beilsten-Edmands, V.; Cannone, G.; Graham, S.; Robinson, C. V.; Spagnolo, L.; White, M. F. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell* **2013**, *52* (1), 124−34.
(39) Staals, R. H. J.; Agari, Y.; Maki-Yonekura, S.; Zhu, Y.; Taylor, D. W.; van Duijn, E.; Barendregt, A.; Vlot, M.; Koehorst, J. J.; Sakamoto, K.; Masuda, A.; Dohmae, N.; Schaap, P. J.; Doudna, J. A.; Heck, A. J. R.; Yonekura, K.; van der Oost, J.; Shinkai, A. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. *Mol. Cell* **2013**, *52* (1), 135−145.
(40) Tamulaitis, G.; Kazlauskiene, M.; Manakova, E.; Venclovas, C.; Nwokeoji, A. O.; Dickman, M. J.; Horvath, P.; Siksnys, V. Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. *Mol. Cell* **2014**, *56* (4), 506−17.

https://dx.doi.org/10.1021/jasms.9b00011
J. Am. Soc. Mass Spectrom. XXXX, XXX, XXX−XXX