Molecular basis for cA6 synthesis by a type III-A CRISPR–Cas enzyme and its conversion to cA4 production

Hemant N. Goswami, Fozieh Ahmadizadeh, Bing Wang, Doreen Addo-Yobo, Yu Zhao, A. Carl Whittington, Huan He, Michael P. Terns and Hong Li

1Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA
2Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA
3Department of Biological Sciences, Florida State University, Tallahassee, FL 32306, USA
4Biochemistry and Molecular Biology, Genetics and Microbiology, University of Georgia, Athens, GA 30602, USA

To whom correspondence should be addressed. Tel: +1 850 644 6785; Email: hong.li@fsu.edu

Abstract

The type III-A (Csm) CRISPR–Cas systems are multi-subunit and multipronged prokaryotic enzymes in guarding the hosts against viral invaders. Beyond cleaving activator RNA transcripts, Csm confers two additional activities: shredding single-stranded DNA and synthesizing cyclic oligoadenylates (cOAs) by the Cas10 subunit. Known Cas10 enzymes exhibit a fascinating diversity in cOA production. Three major forms—cA2, cA6 and cA8 have been identified, each with the potential to trigger unique downstream effects. Whereas the mechanism for cOA-dependent activation is well characterized, the molecular basis for synthesizing different cOA isoforms remains unclear. Here, we present structural characterization of a cA6-producing Csm complex during its activation by an activator RNA. Analysis of the captured intermediates of cA6 synthesis suggests a 3′-to-5′ nucleotidylic transferring process. Three primary adenine binding sites can be identified along the chain elongation path, including a unique tyrosine–threonine dyad found only in the cA6-producing Cas10. Consistently, disrupting the tyrosine–threonine dyad specifically impaired cA6 production while promoting cA4 production. These findings suggest that Cas10 utilizes a unique enzymatic mechanism for forming the phosphodiester bond and has evolved distinct strategies to regulate the cOA chain length.

Graphical abstract

Introduction

In the ever-evolving arms race between prokaryotes and invasive foreign nucleic acids, CRISPR–Cas (clustered regularly interspaced short palindromic repeats and their associated genes) adaptive immune systems play a pivotal role in safeguarding microbial genomes (1–3). The type III-A CRISPR–Cas (Csm) systems, characterized by multiple subunit effector assemblies and activated by invading RNA, epitomize CRISPR–Cas immunity by orchestrating a cascade of transcription-dependent enzymatic activities that effectively...
neutralize viral invaders (4–9). Upon binding of the foreign transcript (target RNA) through base pairing with the guide region of the CRISPR RNA (crRNA), the Csm complex cuts the target RNA at three or four evenly spaced sites. The binding of target RNA further triggers the activation of the HD (histidine-aspartate) nuclease domain and cyclase domain within the Cas10 subunit of Csm, leading to degradation of single-stranded DNA as well as production of cyclic adeny- lates (cA_n, where n = 3, 4, 5 or 6) of variable ring size from ATP, respectively (6,10–12). The COA molecules produced by the activated Csm subsequently elicit secondary immune responses (13–15). Given the destructive potential, the immune responses conveyed by Csm are tightly regulated through an 8-nt sequence at the 3’-end of the target RNA called protospacer flanking sequence (PFS). The RNA containing PFS complementary to the 8-nt tag of the crRNA (5’-tag) is recognized as self, while those containing noncomplementary PFS as foreign (11,12). Interestingly, the immune responses are reset as the target RNA is cleaved, which removes the driving force for activation (11,12).

While the remarkable RNA-stimulated COA production has been identified years ago (11,12), the enzymatic process and the associated regulation for producing specific types of COA molecules remain poorly understood. Each Csm complex consists of five proteins, Csm1 (Cas10), Csm2 (Cas11), Csm3 (Cas7), Csm4 (Cas5) and Csm5 (Cas7), and crRNA, among which the Csm1 subunit, or Cas10, is responsible for the synthesis of COAs via a pair of RNA-dependent polymerase domains (Palm1 and Palm2). Various Csm systems studied to date have been shown to produce a range of COAs, from cA3 to cA6, both in vitro and in vivo (11,16–18). These COA secondary messengers bind to and regulate the function of the downstream effector proteins in a COA-specific manner. For instance, cA4 and cA6 activate the Csm6/Csx1 family of enzymes by binding to their CARF domain, while cA3 activates NucC and SAVED-CHAT enzymes (17–22). Although there have been extensive studies on type-III CRISPR-Cas systems, the differences among the various Csm1 subunits that produce the distinct COAs are not readily clear to explain the specificities.

Previous mechanistic studies of COA synthesis have been focused on the production of cA6. By using inactive Csm1-containing Csm complexes or the Csm1–Csm4 subcomplex incubated with linear or cyclic oligoadenylate molecules, cryo-electron microscopy (cryo-EM) and crystallographic structures revealed two primary adenosine binding sites localized within the two Palm domains of Csm1, respectively (23). At the site associated with the Palm1 domain, an intact ATP (acceptor, ATP1) is bound with its adenine ring being specifically recognized by a tyrosine and a serine residue and its triphosphate group by aspartate residues (23,24). Similarly, at the site associated with the Palm2 domain, another adenosine (donor, ATP2) is accommodated, also being recognized by a tyrosine (or phenylalanine) and a serine residue (24). The observed arrangement of adenosine binding sites led to an intriguing hypothesis regarding successive formation of the phosphodiester bond (23,24). It is believed that the 3’-hydroxyl oxygen of the acceptor attacks the scissile phosphate of the donor, leading to the first linear adenosine dimer (pppA2) that is then attacked by a third incoming ATP at the acceptor site to form a linear adenosine trimer (pppA3) and so on. The final attack takes place in cis between the 3’-hydroxyl of the terminal adenosine and the scissile phosphate, leading to the cyclized product cA6 (23). Similar studies on cA6 synthesis are, however, not available, leaving the questions about its catalytic process and the control of chain length unanswered.

To fill this gap, we employed cryo-EM and mutagenesis approaches to study the cA6 synthesis process of Lactococcus lactis Csm (LiCsm) that is known to produce cA6 (25). We determined cryo-EM structures of the active LiCsm enzyme in the presence of an activating target RNA and ATP. We observed structural features both shared with cA4-producing Csm and unique to cA6-producing Csm, leading to a testable hypothesis for the regulation of the chain length. Mutations of both the shared and unique sites disrupted cA6 synthesis and, interestingly, converted LiCsm from cA6 to cA4 synthesis.

Materials and methods

Cloning

The pACYC LiCsm effector module plasmid encoding Cas6, Csm1–6 and CRISPR loci was as described previously (25). The desired mutations, including Csm1 Y368A and T542A, were introduced by Q5 site-directed mutagenesis (New England Biolabs) (Supplementary Table S4). All constructs were verified by sequencing primers (Eurofins Genomics).

Protein expression and purification

The LiCsm effector complexes were all expressed and purified as described previously (25). Briefly, the all-in-one pACYC plasmid was transformed into the Escherichia coli NiCo21(DE3) strain (New England Biolabs) and the cells were grown to log phase before induction by addition of 0.3 mM isopropyl β-D-1-thiogalactopyranoside. The N-terminal His6-tag on LiCsm2 enabled isolation of LiCsm ribonuclease protein using Ni-NTA affinity chromatography. The Ni-NTA elution pools were loaded on to a size-exclusion column equilibrated with the storage buffer that contains 20 mM Hepes (pH 7.5), 200 mM NaCl, 5 mM MgCl2 and 14 mM 2-mercaptoethanol. Each LiCsm complex was concentrated to ~55 μM and stored at −80°C. The same protocol was followed for purifying LiCsm mutants.

In vitro COA synthesis assay

High-performance liquid chromatography (HPLC) was employed to analyze COA synthesis products. The COA synthesis was conducted by incubating a 250 nM Csm–crRNA binary complex containing the Csm3 D30A mutant (hereafter referred to as LiCsm(Csm3)1, 50 nM target RNA, 1 mM ATP and 10 mM MgCl2 in a reaction buffer (33 mM Tris acetate, pH 7.6, 66 mM potassium acetate) in a 250 μl volume at 37°C for 2 h, 4 h, 9 h and overnight. The reaction was stopped by heat inactivation at 95°C for 10 min and the denatured Csm protein subunits were removed by centrifugation. The final products and control samples were analyzed using an HPLC system (Shimadzu Prominence LC-20) fitted with a SunFire C18 (4.6 mm × 150 mm, 3.5 μm particle size). Sample injections of 5 μl volume were used, employing a linear gradient method with eluent A (20 mM ammonium bicarbonate) and eluent B (100% acetonitrile) at a flow rate of 0.3 ml/min over a 22-min duration. The gradient conditions were as follows: from 0 to 12 min, a gradient of 2–30% B; from 12.1 to 16 min, a transition to 95% B for washing; from 16.1 to 17.0 min, a constant 95% B concentration; and finally, from 17.1 to 22.0 min, a return to 2% B for equilibration. The wavelength used for the detection of COA peaks was 250 nm and the data were acquired using Shimadzu LabSolutions software.

Downloaded from https://academic.oup.com/nar/article/doi/10.1093/nar/gkae603/7710915 by guest on 19 July 2024
for detection of products was 254 nm. The supernatants derived from the LICsm mutants underwent analysis using the identical protocol.

**In vitro RNA cleavage assay**

The cleavage assays were carried out with 500 nM target RNA, 100 nM CARF-HEPN nuclease, 500 nM cA4, 500 nM cA6 and 1 μl of synthesis products of LICsmCsm3 and its cOA binding site mutants. The reaction buffer used for the TaqCxs1-based reaction contained 20 mM MES (pH 6.0) and 5 mM MnCl2, and the reaction buffer used for the LICsm6-based reaction contained 33 mM Tris acetate (pH 7.6) and 66 mM potassium acetate. The TaqCxs1 reactions were incubated for 30 min at 60°C and the LICsm6 reactions were incubated for 30 min at 37°C. All reactions were quenched by adding an equal volume of 2× RNA loading dye (95% formamide, 0.02% sodium dodecyl sulfate, 0.02% bromophenol blue, 0.01% xylene cyanol, 1 mM ethylenediaminetetraacetic acid). The samples were heated for 5 min at 95°C and resolved on 15% polyacrylamide gels (40% acrylamide, 7 M urea, 1× TBE buffer) in 1× TBE running buffer. Nucleic acid bands were stained using SYBR Gold II Nucleic Acid Gel Stain (Invitrogen) and visualized using a Bio-Rad ChemiDoc MP scanner.

**Cryo-EM sample preparation**

The wild-type LICsm–crRNA sample (3 mg/ml) was incubated with 33 μM target RNA and 3 mM ATP at 4°C for 2 h and at 37°C for 5 and 10 min (Supplementary Table S2). Four microliters of the target-bound LICsm complex was applied to glow-discharged UltrAuFoil 300 mesh R1.2/1.3 grids (Quantifoil), blotted for 3 s at 88% humidity and flash-frozen in liquid ethane using FEI Vitrobot Mark IV. The grids were stored in liquid nitrogen before being used for imaging.

**EM data collection, processing and 3D reconstruction**

The ice-embedded LICsm complex samples were collected on an FEI Titan Krios electron microscope equipped with a Gatan Bioquantum K3 direct electron detector (Thermo Fisher Scientific) with the Leginon software for automatic data acquisition in a counting mode (26). All images were collected at nominal 81 000× magnification with a pixel size of 1.074 Å/pixel and a total dose of 60 e−/Å² at −1.0 to −2.5 μm defocus.

Images were processed through both RELION and cryoSPARC (27,28). The motion correction was performed in RELION-4.0 with the UCSF MotionCor2.5 (29). The good micrographs were selected and uploaded to cryoSPARC for subsequent processing. The contrast transfer function (CTF) estimation was conducted through path CTF estimation (30). Particles were initially auto-picked through a blob method and the 2D template was generated and used as a template for complete particle picking. Bad particles were eliminated by several rounds of 2D classification and the remaining good particles were further cleaned and classified by 3D classification. The map from each class was reconstructed by nonhomogeneous reconstruction. The resolution was estimated using the gold-standard Fourier shell correlation plot at the value of 0.143. Local resolution was estimated using local resolution estimation.

A total of 11,658 images were collected for the LICsm complex in a movie mode. Images showing bad ice, astigmatism, drift and poor sample quality were rejected resulting in 9774 images for further processing and particle picking, which resulted in a total of 5,936,523 particles. Several rounds of 2D classification led to 3,039,493 particles with good quality. Good particles/classes were further cleaned in multiclass ab initio and subsequent heterogeneous refinement, which led to three good classes and a total of 2,288,890 particles. A consensus map was generated through nonuniform refinement resulting in an overall resolution of 3.57 Å. Focused 3D classification was conducted with a mask on the Csm1–Csm4 subcomplex generating 10 classes, each of which was refined by nonuniform refinement. The classes with similar features were combined into two different classes. Each class was reconstructed with nonuniform refinement with CTF refinement on the fly, which generated a map of class 1 at a resolution of 2.58 Å and a map of class 2 at a resolution of 2.79 Å.

**Model building and refinement**

The RELION-4.0 maps were used in model building and refinement. The model for the post-cleavage complex (class 1) was first built from the model of 6XN5. A refined model of the post-cleavage complex was used to build the pre-cleavage complex (class 2). ATP and ppA3 were manually built. Real-space refinement was carried out for both models with PHENIX (31). Iterative rounds of real-space refinement and manual building in COOT led to models with satisfactory stereochemistry parameters (32).

**Results**

**Structural difference between RNA pre- and post-cleavage states**

We incubated the in vivo constituted active LICsm with a cognate target RNA and ATP at several temperatures for various time periods under a reactive condition before making cryo-EM grids. We collected and analyzed data for the 5-min (37°C), 10-min (37°C) and 120-min (4°C) reacted complexes and surprisingly found no significant difference in the final reconstructed structures among these different reaction time points, likely owing to the fast dynamics of the reactions under the conditions used. Therefore, we focused on the higher quality structures obtained from the 120-min reaction time point for subsequent analysis (Figure 1, Supplementary Table S1 and Supplementary Figure S1).

The first major structure (class 1, 42%), refined at an overall resolution of 2.58 Å, has a partially cleaved target RNA (+7* to +29*) and a well-resolved Csm1 (Figure 1A). The density for the target RNA abruptly stops for 14 of the 3′-end nucleotides (−8* to +6*) beginning from the Csm3-mediated cleavage site closest to PFS (Figure 1A). The second major structure (class 2, 26%), refined at an overall resolution of 2.79 Å, contains the fully base-paired target RNA to the crRNA and a partially disordered Csm1 (Figure 1B). Neither reconstructed class captured the PFS region of the target RNA (−8 to −1). Given that LICsm used in making cryo-EM specimens is active, class 1 likely represents a target post-cleavage and thus an inactive state, whereas class 2 represents a target pre-cleavage and thus an active state of LICsm. In class 1 where the density is present throughout Csm1, we observed and modeled the acceptor site ATP and the donor site ppA3 (Figure 2). Though the density for Csm1 is weak in class 2, evidence for the presence of the acceptor site ATP and donor site
Figure 1. Overview of the active LfCsm ternary complex and cryo-EM structures. (A) Genome organization of the LfCsm components. (B) Schematic representation of the crRNA and target RNA duplex used in the study. PFS denotes protospacer flanking sequence. The crRNA and target RNA are colored in blue and green, respectively. The light green colored nucleotides depict those not modeled due to insufficient or absent density in class 1 LfCsm. The black triangle indicates the target RNA cleavage sites within the modeled region. (C) Cryo-EM reconstruction of the LfCsm complex (class 1) bound to the post-cleavage target RNA with an ordered Csm1 subunit (left) and the corresponding atomic model (center); the zoom-in view of ATP and pppA3 product in the stick representation. (D) Cryo-EM reconstruction of the LfCsm complex (class 2) bound to the pre-cleavage target RNA with a disordered Csm1 subunit (left) and the corresponding atomic model (center); superimposition of the crRNA–target RNA duplex of class 1 and class 2 and the surrounding loops of the four Csm3 subunits (class 1 is colored in gray and class 2 is colored in blue, green or wheat) (right). In the Cas10 subunit, the bound ATP and pppA3 product are shown as surfaces and colored in yellow; the crRNA and target RNA are shown as surfaces in the atomic model and colored in blue and green, respectively.
ATP exists and that Csm1 could be placed as a rigid body from the refined coordinate in class 1 (Supplementary Figure S1).

The captured post-target cleavage state is consistent with the previous biochemical results from type III systems that cleavage of target RNA begins at the 3’-end closest to PFS and that the 3’ cleavage alone or mismatches in the PFS proximal region are sufficient to deactivate the activities of Cas10 (33,34). Consistently, the class 1 structure largely resembles that of the previously observed apo LIcsm structure that represents an inactive state (PDB ID: 6XN5), especially with respect to the well-ordered Csm1 (25). Interestingly, class 2, in which target remains intact, shares structural features with the cleavage-deficient LIcsm bound with a cognate RNA in the absence of ATP (Supplementary Figure S2), especially the fact that Csm1 is partially disordered (25). The difference in Csm1 structural stability between the inactive class 1 and the active class 2 structure suggests an enhanced motion of Csm1 during activation, consistent with the results from a previous single-molecule study (35).

Identification of cA6-specific binding sites
In class 1 where Csm1 is well ordered, three adenosine nucleotides could be confidently placed into the densities (Figure 2); all are at the interface between the Palm1 and the Palm2 domain. An intact ATP is situated at the cleft primarily formed by Palm1 (acceptor, ATP1), whereas an ATP is at the cleft primarily formed by Palm2 (donor, ATP2) (Figure 2B). These two sites are similar to those previously captured in cA4-producing Cas10 structures suggesting that they are shared between the two types of Csm1 enzymes (23,24,36). Also similar to the cA4-producing Csm1 sites, Csm1-coordinated Mg2+ ions are observed to stabilize the phosphates of the bound ATP molecules. A single Mg2+ ion that stabilizes the 3’ and 2’ phosphates of ATP1 is observed to coordinate with the carboxylate of Asp253 and possibly water molecules. Two Mg2+ ions stabilize ATP2 and are coordinated with the carboxylates of the conserved Asp520 and Asp577, as well as the carbonyl of Val521. The adenine base of ATP1 is specifically anchored by forming hydrogen bonds with the con-
Figure 3. Sequence comparison among cA4- and cA6-producing Csm proteins. Representative sequences from >5600 LlCsm1 homologs are aligned. Residues observed in interaction with adenosine bases of cA4 intermediates are shown and highlighted. Residues belonging to Palm1 are colored in purple and those belonging to Palm2 are colored in green. Species are clearly divided into two clades by the presence and absence of the third tyrosine/serine (threonine) pair and are colored in orange (cA4-producing) and blue (cA6-producing), respectively. The two species with partially conserved pair are colored in black. Top inset shows superimposed ATP/cOA intermediate structures between LlCsm (yellow) and the cA6-specific StCsm (gray, PDB: 6iG0). Bottom inset shows superimposed ATP/cOA intermediate structures between LlCsm (yellow) and the cA4-specific ToCsm (gray, PDB: 6O75).

served Ser277 and Asp281 of Palm1 and aromatic stacking with Tyr572 of Palm2 (Figure 2C). The adenine base of ATP2 is similarly stabilized by a Tyr/Ser pair from both Palm domains, Ser456 of Palm2 and Tyr304 of Palm1, respectively. The cross-stabilization of the two primary ATP sites by both Palm domains signifies the unique architecture of Cas10 proteins that support nontemplated phosphodiester bond synthesis.

In contrast to cA4-producing Cas10, the LlCsm–Cas10 structure reveals an additional adenosine binding site that shares the similar cross-domain characteristics of the other two sites. The density clearly revealed the third adenosine residue (hereafter AMP1) covalently linked to the donor ATP2 (Figure 2), likely as a reaction product of cA6 synthesis during sample preparation. Strikingly, the adenine base of AMP1 protrudes into a pocket to interact with Tyr368 and Thr542, of Palm1 and Palm2, respectively, in an analogous manner to the donor and acceptor ATP. The adenine base of AMP1stacks on the phenyl ring of Tyr368 while forming a hydrogen bond with the hydroxyl group of Thr542.

To learn whether the identified adenosine-interacting residues are cOA-specific among known Csm1 sequences, we analyzed protein sequences of the LlCsm1 homologs. Among the >5000 highly conserved Csm1 sequences ($e$-value <2e−180), we chose one representative from each genus for multiple sequence analysis (Figure 3 and Supplementary Figure S3). Strikingly, based on the third Tyr/Thr pair (Tyr368 and Thr542), the sequences diverge into two clades that coincide with experimentally identified cA6- and cA4-producing Cas10 (Figure 3). In the cA4-producing clade that includes the known Thermococcus omnarius and Thermus thermophilus Cas10, Tyr368 and Thr542 are clearly absent, while the Tyr/Ser pairs critical to ATP1 and ATP2 binding are well conserved (Figure 3). In contrast, in the cA6-producing clade, all three Tyr/Ser(Thr) pairs are conserved (Figure 3). Furthermore, structural superimposition of LlCsm–Cas10 with ppA3-bound ToCsm (cA4-producing) revealed a drastically different trajectory of the third adenosine (AMP1) (Figure 3B), whereas with ATP-bound StCsm1 (cA6-producing) revealed conserved location of the third Tyr/Ser(Thr) pair (Figure 3B) (23,24).

The conserved donor and acceptor ATP binding between the cA6- and cA4-producing Csm1 suggests a conserved catalytic process for phosphodiester bonds. As proposed for the cA4-producing ToCsm1, the 3′-OH of ATP1 could launch an in-line nucleophilic attack on the α-phosphate of the ATP2 to form a 3′−5′ phosphodiester linkage. Asp577 that is coordinated with Mg2+ of ATP1 and as part of the hallmark GGDD motif likely activates the 3′-OH group (Figure 2C). However, similar to what was observed for the cA4-producing ToCsm1, the distance between the 3′-OH and the α-phosphate is too far for the nucleophilic attack (~5.5 Å), signifying that a dynamic motion between Palm1 and Palm2 is required for initiating the catalytic step.

**Impaired cA6-specific binding site promotes cA4 production**

To assess the functional contribution of the structurally identified cA6 binding residues, we performed mutagenesis of the
Cas10 subunit on the LICsm$^{dCsm3}$ complex that lacks target RNA cleavage activity and analyzed the COA synthesis products by HPLC and mass spectrometry (MS). Notably, the equivalent residues stabilizing ATP1 and ATP2 have been studied previously in other Csm systems (23,24). We thus focused on the newly identified Tyr368/Thr542 that stabilizes AMP1. We reasoned that disruption of Tyr368/Thr542 may impair cA$_{6}$, whereas disruption of other Tyr/Ser(Thr) pairs may reduce overall COA synthesis.

The HPLC and MS results suggest that the synthesis products by the LICsm$^{dCsm3}$ are dominated by the presence of linear pppA$_{3}$ and cA$_{6}$ early in reaction (Figure 4, Supplementary Figures S4 and S5, and Supplementary Table S3). There are cA$_{4}$ molecules made after overnight incubation of the LICsm$^{dCsm3}$ with the target RNA, suggesting that LICsm can also produce some cA$_{4}$ under certain reaction conditions such as with molar excess ATP and the target RNA. In contrast, under the same conditions and reaction times, the Thr542 to alanine mutant (T542A) produced cA$_{4}$ much earlier in reaction time with a similar production of cA$_{6}$ (Figure 4 and Supplementary Figures S4 and S5). Strikingly, the Tyr368 and Thr542 to alanine double mutation (Y368A/T542A) produced almost exclusively cA$_{6}$ with minimal synthesis of cA$_{4}$ (Figure 4 and Supplementary Figures S4 and S5), supporting their specific roles in cA$_{6}$ synthesis.

Interestingly, mutation of the ATP2-interacting residue Tyr304 to alanine (Y304A) reduced overnight production of cA$_{4}$ while preserving that of cA$_{6}$ (Figure 4), which reinforces its functional role in overall COA production.

To further demonstrate the functional switch from cA$_{6}$ to cA$_{4}$ production by the T542A/Y368A LICsm variant, we used its COA synthesis products to stimulate the RNA degradation activity of two CARF-HEPN nucleases with known cOA preference. Strikingly, whereas the LICsm$^{dCsm3}$ synthesized products from 2-h reaction did not activate the cA$_{4}$-specific HEPN nuclease, Csx1, the T542A/Y368A (TY) variant elicited robust RNase activity (Figure 5 and Supplementary Figure S6). Both the LICsm$^{dCsm3}$ and its TY variant activated the RNase activity of Csm6, a cA$_{6}$-specific HEPN nuclease, suggesting that the T542A/Y368A variant still retains a level of cA$_{6}$ synthesis despite a significant cA$_{4}$ production (Figure 5), consistent with the HPLC and MS analysis results (Figure 4 and Supplementary Figure S5).

**Discussion**

We employed structural and mutagenesis methods in unveiling previously unknown specificity elements in the synthesis of COA molecules by Csm. The use of an active cA$_{6}$-producing Csm enzyme allowed capturing the cA$_{6}$ synthesis intermedi-
ates, thereby identifying the key Csm1 motifs that control the chain length.

We showed that there are three primary anchoring sites, characterized by Tyr/Ser(Thr) pairs, for adenosine on the Cas10 subunit, two of which are shared with those on cA₄-producing Cas10, while the third pair is unique to cA₅-producing Cas10. Remarkably, by mutating the third Tyr/Ser(Thr) pair, the cA₅-producing Csm complex can be switched to favor cA₄ production. Notably, each Tyr/Ser(Thr) pair that anchors the adenine ring is distributed across the Palm1 and Palm2 domains (Figures 2B and 3A). The composite binding site for the elongating polyadenylate suggests that domain motions between Palm1 and Palm2 are critical to the synthesis of cOA. Conversely, the binding and synthesis of cOA may impact the dynamics of the two domains and the associated Cas10 domains such as the C-terminal domain and the HD domain. We believe that this mechanism is in part responsible for the RNA-activated Csm ancillary activities, which is consistent with the increased flexibility of the Cas10 subunit upon binding of the cognate RNA (25). In type III-B systems, the change in motion between the two Palm domains by stalk-loop-mediated interaction with zinc fingers can potentially regulate cyclase activity (37). Examples of regulating cyclic nucleotide synthesis through dual Palm motions are found in other nucleotide cyclases such as the zinc-triggered diguanylate synthesis (38). The diguanylate synthesis enzymes form a homodimer, placing two Palm domains close together with one GTP anchored on each (Supplementary Figure S7). Binding of the zinc ions to the regulatory domain fused to the Palm domain regulates the relative distance between the two Palm domains, thereby activating diguanylate synthesis (38).

The captured pppAp₃ and ATP on LICsm during cA₅ synthesis suggest that Cas10 catalyzes nontemplated 3’ → 5’ addition of nucleotides, unlike the conventional 5’ → 3’ addition by all nucleotide polymerases, including DNA and RNA polymerases, reverse transcriptase and telomerase (Figure 6). The free 3’-OH of the incoming ATP at the acceptor site (ATP1) attacks the α-phosphate of the ATP or pppAp₄₋₁ at the donor site (ATP2), leading to the formation of pppAp₅ (n = 2, 3, 4, 5). The process repeats until pppAp₅ is formed, which is then cyclized by the attack of 3’-OH of the sixth adenosine to the 5’-triphosphate. The sequential addition of adenosine is consistent with the detection of cA₄ and cA₅ oligoadenylates by MS (Supplementary Figure S5). The highly unusual 3’→5’ nucleotidyl transfersase have been observed previously, including the evolutionarily conserved tRNAHis guanylyltransferase (Thg1) that adds a single guanine to the 3’-end of tRNAHis (39). Unlike Cas10, however, Thg1 employs the free 3’-OH of GTP to attack the α-phosphate of the 5’-adenylylated-tRNAHis intermediate and a water molecule to remove the
pyrophosphate from the triphosphorylated pppG−1−tRNA3His. Significantly, both Thg1 and Cas10 share the same catalytic architecture with the canonical polymerases, reflecting the similar chemical steps involved in their respective nucleotidyl transferring reactions.

The proposed sequential synthesis mechanism aligns well with the asymmetric binding—the elongating chain occupies Palm2, while the incoming ATP binds to Palm1. This contrasts with a parallel synthesis model, where two pppA3 molecules would need to bind symmetrically to form cA6. Prior research on the ToCsm complex bound with synthetic pppA3 also revealed a single pppA3 molecule occupying Palm2, lending further support to the sequential synthesis mechanism (23). Interestingly, our cryo-EM structures, obtained from three independent data sets under various conditions, consistently show a pppA3 intermediate but lack the evidence for longer intermediates beyond pppA3. This might be due to the inherent flexibility of the longer chains, making them difficult to visualize. Alternatively, the conversion of pppA3 to pppA4 could be the rate-limiting step in the reaction pathway. In line with this idea, our HPLC and MS results indicate that LiCsmΔCas3primarily produces pppA3 and cA6, suggesting that the formation of the pppA3−1 intermediate is the slowest step in cA6 synthesis.

While our conclusions fill a gap in understanding the enzymatic mechanism in the synthesis of cA6 molecules, they also offer potential advancements in practical applications of the LiCsm system. LiCsm has previously been repurposed for RNA detection through its cA6 production that activates an ancillary cA4-specific CARF–HEPN nuclease (40). The production of cA4 by the T542A/Y368A variant opens doors to linking RNA sensing by LiCsm to both cA4− and cA6−specific effector enzymes, thereby broadening the utilization of the tandem enzyme strategy for diagnosis.

Data availability

The atomic coordinates of the cryo-EM structures of class 1 and class 2 have been deposited in the Protein Data Bank under the identifiers 9ASH and 9ASI, and in the Electron Microscopy Data Bank under the entries EMD-43814 and EMD-43815, respectively.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We are grateful for indispensable support from Xiaofeng Fu for assistance in cryo-EM data collection, Huan He for assistance in MS analysis, Gwimoon Seo for assistance in protein expression and Steven Miller for DNA sequencing analysis throughout this study. We acknowledge the utilization of instruments at the BSIR, including the Titan Krios, by National Institutes of Health (NIH) grant S10 RR023080, the Vitrobot and Solaris Plasma Cleaner, supported by S10 RR024564, and the BioQuantum/K3, acquired through U24 GM116788 funding. We are also thankful to SECM4 (The Southeastern Center for Microscopy of Macromolecular Machines) for screening and data collection through NIH grant R24 GM145964.
Author contributions: H.N.G. and H.L. designed the experiments; F.A. expressed and purified LiCsm complexes; Y.Z. provided assistance in grid preparation; B.W. and H.N.G. performed cryo-EM analysis; H.N.G. and A.C.W. performed COA synthesis assay by HPLC; D.A.-Y. performed RNA cleavage assays; H.H. performed MS experiments; H.N.G. and B.W. made the figures; H.N.G., M.P.T. and H.L. wrote the manuscript. All authors edited the manuscript and provided insightful comments.

Funding
National Institutes of Health [R35 GM152081 to H.L., R35 GM118160 to M.P.T.]. Funding for open access charge: National Institutes of Health [R35 GM152081].

Conflict of interest statement
None declared.

References


