Structural basis of cyclic oligoadenylate binding to the transcription factor Csa3 outlines cross talk between type III and type I CRISPR systems

Received for publication, October 16, 2021, and in revised form, January 11, 2022 Published, Papers in Press, January 14, 2022, https://doi.org/10.1016/j.jbc.2022.101591

Pengjun Xia 1, Anirudha Dutta 2, Kushol Gupta 3, Mona Batish 1,2, and Vijay Parashar 1,2,*

From the 1Department of Biological Sciences, and 2Department of Medical and Molecular Sciences, University of Delaware, Newark, Delaware, USA; 3The Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Edited by Wolfgang Peti

RNA interference by type III CRISPR systems results in the synthesis of cyclic oligoadenylate (cOA) second messengers, which are known to bind and regulate various CARF domain–containing nuclease receptors. The CARF domain–containing Csa3 family of transcriptional factors associated with the DNA-targeting type I CRISPR systems regulate expression of various CRISPR and DNA repair genes in many prokaryotes. In this study, we extend the known receptor repertoire of cOA messengers to include transcriptional factors by demonstrating specific binding of cyclic tetra-adenylate (cA4) to Saccharolobus solfataricus Csa3 (Csa3sso). Our 2.0-Å resolution X-ray crystal structure of cA4-bound full-length Csa3sso reveals the binding of its CARF domain to an elongated conformation of cA4. Using cA4 binding affinity analyses of Csa3sso mutants targeting the observed Csa3sso–cA4 structural interface, we identified a Csa3-specific cA4 binding motif distinct from a more widely conserved cOA-binding CARF motif. Using a rational surface engineering approach, we increased the cA4 binding affinity of Csa3sso up to ~145-fold over the wildtype, which has potential applications for future second messenger-driven CRISPR gene expression and editing systems. Our in-solution Csa3sso–cA4 structural analysis identified cA4-induced allosteric and asymmetric conformational rearrangement of its C-terminal winged helix-turn-helix effector domains, which could potentially be incompatible to DNA binding. However, specific in vitro binding of the purified Csa3sso to its putative promoter (Pcasa) was found to be cA4 independent, suggesting a complex mode of Csa3sso regulation. Overall, our results support cA4-and Csa3-mediated cross talk between type III and type I CRISPR systems.

Clustered regularly interspaced short palindromic repeat (CRISPR)-Cas systems mediate prokaryotic immune response to mobile genetic elements (MGEs) (1) and have been identified in ~40% bacterial and ~90% archaeal genomes (2–4). These systems typically contain (i) CRISPR arrays containing alternating identical repeats and unique spacers derived from the MGEs (5, 6) and (ii) cas gene cassettes adjacent to the CRISPR arrays that encode helicases, nucleases, and structural proteins (7).

The hallmark property of the CRISPR-Cas systems is their ability to adapt, process, and interfere with foreign genetic material. This is accomplished through a three-stage process involving (i) spacer acquisition (adaptation), (ii) crRNA production (processing), and (iii) target interference (8). The adaptation stage involves de novo spacer acquisition by a complex formed by two highly conserved nucleases (Cas1 and Cas2) (9, 10). A Cas4 endonuclease further recognizes and removes a short 5′-flanking region of the prespacers called protospacer adjacent motif (PAM) ensuring integration of only mature spacers into the CRISPR array (11). During the crRNA production stage, Cas6 (or an endogenous protein) processes long CRISPR pre-RNAs into short mature crRNAs to contain a region of the extrachromosomal genetic element, a 5′-tag derived from the preceding repeat, and a 3′ end handle from the downstream repeat (12–17). During the final stage of target interference, a ribonucleoprotein (RNP) complex, comprising the mature crRNA and Cas proteins, specifically cleaves foreign nucleic acids via base pairing with the crRNA (18) and by the recognition of the PAM sequence in the target (19–24).

Based on the complexity of the associated RNP complexes, CRISPR systems are classified into two major classes (class 1 and class 2) that contain a total of six different types (types I, III, and IV for class 1 and types II, V, and VI for class 2) (25, 26). The interference RNP complexes of class 2 systems employ a single protein in complex with a crRNA, whereas class 1 systems use multisubunit RNP complexes. Owing to their simplicity and amenability to practical applications such as genome editing, the class 2 systems (containing Cas9, Cas12, and Cas13) are more well studied than class 1 systems (27, 28). The less studied class 1 systems, however, are more primitive, abundant, and widespread in prokaryotes comprising about 90% of all CRISPR-Cas systems (25, 29).

The most widespread type I and type III class 1 systems coexist in many prokaryotic genomes (30–32). Such a coexistence is well represented in genomes from a crenarchaeal
order Sulfolobales, two-thirds of which harbor both the systems (4, 33, 34). For example, *Sulfolobus islandicus* encodes two CRISPR loci, one subtype I-A and two subtype III-B interference modules (34). *Saccharolobus solfataricus*, on the other hand, possesses a more extensive CRISPR system with six different CRISPR loci, which include two type I adaptation modules as well as three type I and four type III interference modules (35, 36).

The type I and III interference complexes exhibit a striking functional diversity (37). Recognition of a PAM and a crRNA complementary sequence in the target DNA by the type I interference complex (known as CRISPR-associated complex for antiviral defense or Cascade) recruits the Cas3 endonuclease for degradation of the nontarget strand (Fig. 1B) (38, 39). By contrast, the type III interference (cmr−csm) complexes recognize a newly transcribed phage RNA in a PAM-independent fashion by base pairing with a seed motif at the 3′ end of the crRNA (4, 30–34). Self-targeting during type III interference is prevented by a mismatch at the 3′ end of the crRNA (Fig. 1A) (40, 41). The interference by cmr−csm complexes involves degradation of the target RNA as well as the nontemplate DNA (Fig. 1A) (32, 42–47). Owing to its PAM-independent functional mode, the cmr−csm complex exhibits a broad target specificity and provides a unique phage survival advantage to the prokaryotic cells coharboring the type I and III systems (48, 49).

Upon recognition of the target RNA, the palm domain of the Cas10 subunit of the cmr−csm complex synthesizes cyclic oligoadenylates (cOAs) containing 3 to 6 adenylate groups (named cA3-cA6) (Fig. 1A), a function that is deactivated upon target RNA cleavage (50–52). cOAs are important second messengers that orchestrate an antiviral response by primarily binding to CRISPR-associated Rossmann fold (CARF) domains in many CRISPR-associated proteins (Fig. 1A) (53). cOA binding regulates the function of various nucleic acid hydrolases such as Csm6/Csx1 family ribonucleases, CRISPR ancillary nuclease 1 (Can1) proteins, and cOA-activated RNase and DNase 1 (Card1 or Can2) proteins. Furthermore, many of the CARF-containing cOA receptors harbor a cOA phosphodiesterase (also called “ring nuclease”) activity that hydrolyzes the cOA ligand into a linearized tetranucleotide (A4 P) and then into two linear A2 > P species (54). The ring nuclease cOA receptors include (i) standalone host nucleases such as CRISPR ring nuclease 1 (Crn1), CRISPR ring nuclease 3 (Crm3), and CRISPR ring nuclease 2 (Crm2); (ii) stand-alone viral nuclease AcrlIII1; and (iii) self-inactivating effector nuclease Csm6 (54). The ring nuclease activity of these receptors is believed to prevent a nonspecific nuclease response post MGE clearance by tightly coupling cOA-mediated response to MGE transcription (54).

Bacteriophage infections drive large global changes in the archaeal transcription. For example, infection by *S. islandicus* rod-shaped virus 2 has been shown to upregulate the expression of approximately one-third of the *S. islandicus* genome (55). Although production of cOAs by type III interference complex could conceivably mediate such transcriptional regulation, type III loci do not encode a transcription factor. The adaptation and interference cassettes of the type I systems, however, encode CRISPR Apn3 (Csa3) family transcription factors Csa3a and Csa3b, respectively (56). The *S. islandicus* Csa3b (Csa3bSis) acts as a transcriptional repressor to genes encoding subtype I-A CRISPR spacer acquisition complex, and subtype I-A target interference complex, as well as a transcriptional activator to genes encoding subtype III-B cmr interference complex (57, 58). *S. islandicus* Csa3a (Csa3aSis), on the other hand, transcriptionally activates expression of CRISPR arrays, subtype I-A adaptation complex, and DNA repair proteins (59, 60). An atomic structure of the apo form of a Csa3a homolog from *S. solfataricus* (Csa3Sso) has been previously reported to harbor an N-terminal CARF and a C-terminal MarR-like winged helix-turn-helix (wHTH) domain (61). The Csa3Sso CARF domain exhibited a dimerization-mediated 2-fold symmetric ligand-binding pocket, which was predicted to bind a four-nucleotide-long RNA (61). Consistent with this, Csa3bSis has recently been shown to bind a linear analog of cA4 (5′CAAAAT′) in a CARF domain–dependent way (57). However, despite functional significance of Csa3 transcription factors, their ligand specificity and the structural basis of ligand binding has not been reported. Here, we identify cA4 as the cognate ligand of Csa3Sso, and show that Csa3Sso lacks ring nuclease activity in vitro. We determine a 2.0-Å crystal structure of Csa3Sso bound to cA4 and identify Csa3Sso residues important for cA4 binding. Complementary SAXS analysis indicates a cA4-induced conformational change in preformed Csa3Sso dimers, suggesting that allosteric changes within the Csa3Sso dimer may regulate Csa3-mediated signaling.

**Results**

### Csa3Sso specifically binds cyclic oligoadenylate 4

Based on previous identification of a 2-fold symmetric ligand-binding pocket at the Csa3Sso (a Csa3a homolog, KEGG accession number Sso1445) dimer interface (61), and in vitro binding of a cA4 analog to Csa3bSis (57), we hypothesized that the Csa3 family of transcription factors from *S. solfataricus* could also be receptors of cOAs. To test specificity of binding of Csa3Sso to cOA nucleotides, we purified an N-terminally His6-tagged fusion of Csa3Sso (His6-Csa3Sso) recombinantly produced in *Escherichia coli* (Fig. 1C) and performed binding affinity analyses of cA3, cA4, cA6, and the linear 5′CAAAAT′ RNA analog using microscale thermophoresis (MST). Despite a nonspecific binding exhibited by high concentrations of all the cOAs (250–1000 μM) (Fig. S1), only cA4 exhibited a specific low micromolar binding to His6-Csa3Sso with an apparent dissociation constant (KD) of 5.8 ± 0.03 μM (Fig. 1D). This is consistent with the previously observed predominance of cA4 among the cOAs produced by *S. solfataricus* Csm complex in vitro (50). Of note, the lack of Csa3Sso binding to cA3, a second messenger also produced by CD-NTases of an alternate CBASS antiviral defense system in bacteria (62), suggests specificity of the Csa3 transcription factors to CRISPR-Cas systems. Furthermore, His6-Csa3Sso binding to cA4 is stronger (8-fold) than the previously reported binding.
of Csa3bSso with the linear “cA4 analog” (5’CAAAA3’, $K_D$ of 46.10 ± 8.14 μM) that was determined using surface plasmon resonance (57). However, both these Csa3 homologs from Sulfolobus show cA4 binding affinities lower than Treponema succinifaciens Card1 ($K_D$ of 15 nM), which is also a cA4 receptor lacking ring nuclease activity (63). Nevertheless,
considering the expected high micromolar concentrations of cA4 in the cell upon infection as discussed below (64), we believe that cA4 is the preferred ligand of Csa3Sso.

The 2.0-Å crystal structure of Csa3Sso bound to cA4

To better understand the structural basis for Csa3Sso ligand specificity, we determined a 2.05-Å X-ray crystal structure of His6-Csa3Sso bound to cA4 (Table 1). The overall Csa3Sso•cA4 structure is very similar to that of the previously determined 1.8 Å apo Csa3Sso structure (PDB ID 2WTE, RMSDs of 0.735 Å over all atoms, Fig. S2)(61). Like the apo Csa3Sso structure, the Csa3Sso dimer is domain swapped with respect to an N-terminal CARF and the C-terminal wHTH DNA-binding domains of the two protomers A and B (Fig. 2). The N-terminal CARF domain in each protomer is composed of six mixed β-strands flanked by four α-helices. The first five β-strands (βN1−βN5) run parallel, whereas the last one (βN6) runs antiparallel and connects the CARF domain to a C-terminal wHTH DNA-binding domain (residues 145−212) through a linker (residues 133−144). The linker is composed of two turns of α-helices followed by four residues in an elongated conformation. The C-terminal wHTH DNA-binding domain contains a right-handed three-helix bundle formed by three helices (αC1−αC3) where αC2 and αC3 comprise the wHTH motif with αC3 (residues 173−186) constituting the DNA recognition helix. The wHTH of Csa3Sso belongs to the widespread MarR-like wHTH fold in which the DNA recognition helix is followed by two β-strands to make a “wing” that follows an α-helix. Accordingly, the HTH wing in the His6-Csa3Sso•cA4 structure is composed of the αC3−αC4 loop that is followed by αC4. Although residues 192 to 196 at the tip of this wing are disordered in protomer A, the electron density for all the protomer B wing residues was observed likely due to their stabilization by contacts with a symmetry-related subunit (Fig. 2A).

The Csa3Sso•cA4 complex exists as a dimer in vitro

Apo Csa3Sso has previously been shown to exist in a dimeric form (61). However, crystal packing analysis of our Csa3Sso•cA4 crystal structure depicted dimer−dimer

---

Figure 2. X-ray crystal structure of Csa3Sso•cA4 complex at 2.05 Å. A, a side view of the X-ray crystal structure of Csa3Sso dimer complexed with cA4 (ball-and-stick model) using its N-terminal CARF domain. Blue and purple cartoons represent two Csa3Sso protomers in the dimer. The Csa3Sso secondary structure elements are labeled with “type” (α or β), followed by “domain” (N or C terminal), and “number” (1−6 for β strands and 1−4 for α-helices). Protomer B elements are labeled with an apostrophe symbol (’) to differentiate them from protomer A elements. αC3 is connected to αC4 by a flexible linker shown as a dashed blue line. The zoomed-in inset on the top shows an expanded side view of cA4 (ball-and-stick model), highlighting the “planar” and “outward facing” (outer) groups in cA4. B, a top view of the Csa3Sso•cA4 structure showing Csa3Sso in the surface representation (blue and purple). To obtain this view of the Csa3Sso•cA4 model, the structure illustrated in A was rotated 90° in the direction indicated by the black arrow. The zoomed-in inset on the top shows an expanded top view of cA4 (ball-and-stick model), with a composite electron density map (2Fo−Fc, contoured at 2.0o) of cA4 in the refined Csa3Sso•cA4 structure. In both A and B insets, groups in cA4 are denoted as A, adenine; R, ribose; P, phosphoryl. To simplify illustration of the bound cA4 in this text, we refer to its AMP moieties as AMP11, AMP12, AMP21, and AMP22; the corresponding adenine rings as A11, A12, A21, and A22; its ribose rings as R11, R12, R21, and R22; and the four phosphoryl groups as P11 (connecting R11 and R21), P22 (connecting R12 and R22), P12 (connecting R11 and R22), and P21 (connecting R12 and R21). All the cA4 atoms had a clash score of <0.7 except for C4 and O4 atoms of R22, which was allowed at 1.15 Å to fit cA4 molecule in the electron density.
interactions that could underlie a higher-order oligomerization. More specifically, a symmetry mate obtained by operation X = 1/2, −Y + 1/2, −Z = 1 to the dimer in the asymmetric unit showed an increase in buried surface area of 2444.5 Å² per dimer of Csa3Sso, using the PISA server (65, 66). Such a large crystallographic interface does not exist in the apo Csa3Sso structure. To assess the possibility of cA4-induced Csa3Sso oligomerization in solution, we performed size-exclusion chromatography. Consistent with prior reports, our Csa3Sso preparations eluted as dimers in gel filtration chromatography (~60 kDa versus globular standards, Fig. 1C). We further analyzed the oligomeric properties of Csa3Sso in the presence and absence of cA4 using sedimentation velocity analytical ultracentrifugation (SV-AUC) (Figs. 3 and S3 and Table S1). In SV-AUC, His6-Csa3Sso (96 μM) appeared as a single peak at 3.5 S20,w with estimated mass (M0) of 52 kDa, consistent with the theoretically calculated S and mass values of 3.8 S and 52.3 kDa from the apo Csa3 structure (PDB 2WTE) (Figs. 3 and S3 and Table S1) (61). The addition of excess amounts of cA4 (Figs. 3 and S3 and Table S1) yielded a very similar distribution as well (M0) of 52 kDa, consistent with the theoretically calculated S and mass values from our Csa3Sso-cA4 structure with cA4 bound (3.86 S20,w and 57.5 kDa). Overall, these data evidence monodisperse dimers of Csa3Sso in solution that persist in the presence and absence of Csa3Sso. Consistent with in-solution dimers observed for Csa3Sso, the asymmetric unit for the Csa3Sso-cA4 complex structure consisted of a dimer of Csa3Sso bound to one molecule of cA4 (Fig. 2B).

Conformation adopted by cA4 in the Csa3Sso-cA4 structure
Comparison of all the cA4-bound structures of CARF domain proteins revealed that cA4 in the Csa3Sso–cA4 complex structure exists in a unique elongated conformation. In this conformation, P1a and P1b (called distal phosphoryl groups) in the cA4 ring are stretched outwards (with a P-to-P distance of 11.8 Å), which brings P2a and P2b (called proximal phosphoryl groups) to a shorter P-to-P distance of 4.4 Å (Fig. 2B inset). The difference of 7.4 Å in these P-to-P distances exemplifies the most extended conformation of cA4 observed among known cA4 receptors (Fig. S4). Furthermore, A2a and A2b adenines adopt a conformation parallel to the plane of cA4 backbone (hereafter called planar adenines), whereas A1a and A1b adenines face outside of this plane pointing away from the binding pocket (hereafter called nonplanar adenines [Fig. 2A inset]). Similarly, 2’ hydroxyls in R1a and R1b, and both unbonded oxygens in P2a face away from the binding pocket (Fig. 2A inset). Three of the phosphoryl groups in cA4 (labeled P1a, P1b, and P2b) are planar, whereas P2a faces away from the binding pocket (hereafter called nonplanar phosphoryl). Furthermore, the 2’ hydroxyl oxygen (O’), phosphorous (P), and 5’ ribosyl oxygen (O”) of R1a, R1b, R2a, and R2b exhibit O’-P”-O” angles of 125°, 141°, 158°, and 94°, respectively (Fig. 4C).

**cA4 recognition by Csa3Sso**
Superimposition of our cA4-bound structure with the apo-Csa3Sso structure (PDB 2WTE) revealed that the cA4-binding pocket is largely preformed upon dimerization of the CARF domain (residues 1–132) (Fig. 2). Owing to symmetry in this pocket, overlapping sets of CARF residues from both the Csa3Sso protomers make equivalent interactions with the two halves of cA4. More specifically, CARF residues from proto- mers A and B of the Csa3Sso dimer (labeled with subscripts A and B in this text) interact almost exclusively with adenine-, ribose-, and terminal phosphoryl groups labeled “a” (for A1a, A2a, R1a, R2a, and P1a) and “b” (for A1b, A2b, R1b, R2b, and P1b), respectively (Fig. 4, A–D). Our structural analysis of the Csa3Sso–cA4 interface as well as sequence alignment the ten archaeal Csa3Sso homologs refined the two previously predicted ligand-binding motifs within the Csa3Sso CARF domain (Figs. 4 and 5, also see Discussion) (61). The first motif (named cA4 binding motif 1) comprises residues 8 to 14 (from the β1–α1 loop and the α1 helix), and the second motif (named cA4 binding motif 2) comprises residues 94 to 99 (located in the β4–α4 loop and α4 helix). Residues from these motifs as well as other regions in Csa3Sso make an extensive network of hydrogen bonding and hydrophobic interactions with different cA4 groups (Fig. 4, A and B) as described below.

*Csa3Sso interactions with cA4 adenine rings*
Extensive hydrophobic interactions from residues in cA4 binding motifs 1 and 2 stabilize all four cA4 adenine rings. The nonplanar A1 adenines (A1a and A1b) dock into shallow hydrophobic pockets formed by Thr13A&B and Phe14A&B (from cA4 binding motif 1), whereas the planar A2 adenines (A2a and A2b) occupy much elaborate and deep hydrophobic pockets. In the planar adenine pocket, one face of the adenine is stabilized by a fully conserved Phe10A&B (cA4 binding motif 2), Pro35A&B, Val39A&B, Thr42A&B, and T13B (Fig. 4, A and B). The hallmark of all the adenine-binding pockets in Csa3Sso are π-stacking interactions of Phe14A&B.
Figure 4. Interactions of cA4 with Csa3Sso-A and B, symmetric interactions of the Csa3Sso protomers A and B with the two halves of cA4. A, a LigPlot+ illustration of Csa3Sso interactions with cA4 showing 2-fold symmetry in the binding pocket. The cA4 and interacting Csa3Sso residues are shown in a ball-and-stick representation with green, purple, and blue sticks representing cA4, promoter A, and promoter B, respectively. To depict the symmetric orientation of cA4 and its interactions with Csa3Sso residues with respect to CARF dimerization interface, the corresponding Csa3Sso protomer regions are shown by blue (protomer A) and purple (protomer B) backgrounds. B, a top view of the Csa3Sso/cA4 complex structure showing interactions of cA4 with CARF domains from the two Csa3Sso protomers. Csa3Sso protomers A and B are depicted in blue surface and purple cartoon representations, respectively. For simplicity of illustration, only the hydrophobic interactions from protomer B (purple residues in ball-and-stick representation) and only the polar interactions from protomer A (blue residues in ball-and-stick representation) are shown. The hydrogen bonds are depicted as solid yellow (B) and dashed green (A) lines with cA4-induced allostery in Csa3.
and Phe10A&B with the nonplanar (A1a and A1b) and planar (A2a and A2b) adenine rings, respectively. More specifically, Phe14A and Phe10B stabilize adenines A1a and A2b by T-shaped π-stacking interactions, and sandwich-type π-stacking interactions from Phe14B and Phe10A engage A1b and A2a, respectively (Fig. 4C). Both Phe14 and Phe10 are significantly conserved in Csa3Sso homologs; Phe14 is conserved in six of the ten Csa3Sso homologs, and Phe10 is conserved as a Phe or Tyr in all ten homologs (Fig. 5). Furthermore, structural superimposition of cA4-bound Csa3Sso with apo Csa3Sso shows linear movement of Phe14A or 90° rotation of Phe14B likely facilitating interactions with cA4 (data not shown). To analyze the contribution of Phe14 and Phe10 to cA4 binding, we mutagenized these residues to alanines and investigated their ability to bind cA4 by MST. The alanine mutants of both Phe14 and Phe10 showed a complete loss of cA4 binding. Figure 5. Structure-guided sequence alignment of Csa3 homologs. The Csa3Sso residues at the Csa3Sso/cA4 structural interface are identified by circles below the alignment. The Csa3Sso residues marked with stars at the bottom of alignment identify cA4 binding motifs 1 and 2 as determined from the Csa3Sso*cA4 structure. Circles filled with green triangles identify conserved (>65% similarity among ten Csa3 homologs) residues at the cA4 interfaces in both the Csa3Sso protomers that showed significant loss or gain of function in the cA4 binding assay (Fig. 4E). Empty circles and those filled with smaller circles represent the interfacial residues present in one or both protomer(s) in the Csa3Sso dimer, respectively, and were not subjected to mutagenesis. Circles filled with red color identify interfacial residues that are conserved but were not subjected to mutagenesis in this study. The numbering is based on the residue positions of the Csa3Sso (gene accession number: Sso1445). The conservation of residues at each position is depicted by the size of the letters in the sequence logo on top of alignment, where the most conserved residues are highlighted by a larger-sized letter and a by a black background. Logo letters colored blue, green, red, and black indicate basic, polar, acidic, and hydrophobic residues, respectively. The secondary structure elements are derived from the Csa3Sso/cA4 structure (in which α-helices are shown as magenta cylinders and β-sheets are shown as yellow arrows). The sequence alignments of the Csa3 homologs were performed using the T-Coffee method (107) and were edited using Geneious Prime software (https://www.geneious.com) and Adobe Illustrator (version 25.3). Each homolog is identified by its accession number and bacterial source.
in vitro highlighting their important role in cA4 stacking (Fig. 4E).

Among the polar interactions of Csa3Sso with the cA4 adenosines, side chains of Asn11A&B hydrogen bond to N7 in the nonplanar adenines (A1a and A1b). Since Asn11 is conserved as an Asn, Asp, or His in most of the Csa3 homologs (Fig. 5), it is conceivable that the side-chain carboxylic oxygen or ring nitrogen atoms in these homologs could instead hydrogen bond to the nonplanar adenines (Fig. 4, A and B). Furthermore, the main-chain oxygen atoms of Met8A&B (cA4 binding motif 1) and Glu122A&B (β5–β6 loop) also hydrogen bond N3 in the planar (A2a and A2b) and nonplanar (A1a and A1b) adenines, respectively (Fig. 4, A and B).

Csa3Sso interactions with phosphate and ribose groups at distal ends of cA4

The P1 phosphoryls (P1a and P1b) and R2 ribosyls (R2a and R2b) at the distal ends of the elongated cA4 are hydrogen bonded by main-chain atoms of many residues in the cA4 binding motifs 1 and 2 (Fig. 4, A and B). More specifically, the main-chain nitrogen atoms of Phe10A&B hydrogen bond with 2’ ribosyl oxygens of R2 groups (R2a and R2b) as well as with P1 phosphoryl oxygens (P1a and P1b). The main-chain nitrogen atoms of Gly9A, Asn11A&B, and Gly96A&B hydrogen bond with the P1 phosphoryl oxygens (P1a and P1b). Among hydrophobic interactions, main-chain atoms of Gly9B, Gly96A&B, and Met95A&B further stabilize the P1 and R2 groups. Of interest, Gly9 and Gly96 are parts of β1–α1 and β4–α4 loops in motifs 1 and 2, respectively, and are completely conserved among Csa3 homologs (Fig. 5). It is therefore possible that these glycines add flexibility to these loops to facilitate binding of the P1 phosphoryl groups in an extended cA4 conformation (Fig. S4). Furthermore, Csa3Sso-Met95 is conserved as a glycine in most of the Csa3 homologs, which may further be adding to the flexibility of the β4–α4 loop to accommodate this cA4 conformation (Fig. 5). To test whether small glycine residues play a role in binding, we mutated Gly96 to a slightly bulkier residue alanine. Indeed, the G96A mutant completely lost the ability to bind cA4 in our MST assay, confirming the requirement of a small residue at this position (Fig. 4E).

Csa3Sso interactions with proximal cA4 phosphoryls

The inward-facing phosphoryl group in the middle of the elongated cA4 (P2b) was found to hydrogen bond with highly conserved Arg98A&B (cA4 binding motif 2) and Glu122A (β5–α5 loop) residues in Csa3Sso. The interactions with Arg98 involve hydrogen bonds to P2b oxygens using either two (for Arg98b) or one (for Arg98a) of its side-chain nitrogen atoms (Fig. 4, A and B). Arg98 is completely conserved in all the Csa3Sso homologs (Fig. 5). Accordingly, our alanine mutant of Arg98 completely lost cA4 binding in vitro (Fig. 4E). Since Arg98 also contributes significantly to dimerization interface of Csa3Sso (buried surface area per Arg = 256.14 Å²), we confirmed that this mutant is still dimeric in solution using SV-AUC (Figs. 3 and S3 and Table S1).

Of interest, the cA4 phosphoryl interacting surface of the Csa3Sso pocket is largely positively charged except for the conserved Glu122 residue that hydrogen bonds with cA4 central phosphoryl (P2b) oxygen via its side-chain carboxyl in the protomer A (Fig. 4, A and B). Such interactions of Glu122 side chains could create an electrostatic repulsion with central phosphorlys (P2a and P2b) likely constraining them close to each other in the elongated cA4 conformation (Fig. 4D). The Glu122A&B side-chain carboxyl also hydrogen bonds with the main-chain nitrogen of Arg98bA in the alternate protomer across the dimerization interface (Fig. 4C). Given the role of Arg98- and Gly96-containing motif 2 in cA4 binding discussed above, we wondered if Glu122 interactions with Arg98 main chain help position these motif 2 residues to affect cA4 binding. We therefore hypothesized that substitution of the Glu122 side chain to an Ala should remove electrostatic Glu122 repulsion to the central phosphoryls and disrupt polar interactions with motif 2, which should facilitate Csa3Sso binding to cA4. Indeed, a mutation of Glu122 to alanine (E122A) dramatically improved the binding affinity of Csa3Sso for cA4 (Kd of 38.3 nM, ~145-fold higher affinity over wild-type) (Fig. 4E). To further dissect whether this cA4 binding gain is due to removal of side chain charge, or to the disruption of interactions with Arg98 main chain, we mutagenized Glu122 to a neutral Gln residue, which should still be able to make polar interactions with Arg98 (and P2b). An E122Q mutation was also found to drastically increase the cA4 binding affinity in vitro (Kd of 44.5 nM, ~125-fold over wildtype) showing a more significant role of electrostatic repulsion by Glu122 as compared with its coordination of Arg98 (Fig. 4E). In summary, these results reveal the mechanistic basis underlying the inability of wildtype Csa3Sso to strongly bind cA4 and further confirm that we have identified a biologically important Csa3 protein surface.

Csa3Sso lacks ring nuclease activity

In the CARF-containing ring nucleases, the motif I is primarily implicated in the catalytic activity (67). Nevertheless, a conserved COA-binding lysine in motif II of cA4 ring nucleases (Lys168 in Sso1393 and Lys106 in Sso2081) (67) was predicted to participate in COA catalysis by stabilizing the transition state (68). Given Arg98A&B coordination with P2b at the end of the β4–α4 loop in Csa3Sso•cA4 structure, we tested cA4 ring nuclease activity of Csa3Sso in two different conditions (Figs. 6 and S5, A and B). A C18 HPLC analysis of the reaction components after removal of Csa3Sso showed no sign of cA4 hydrolysis products (Fig. 6). Consistent with this, P2b exhibits an O’–P–O’ angle of 94° in the Csa3Sso•cA4 structure, which is inconducive for an inline nucleophilic attack by the 2’ ribosyl hydroxyl.

In a few CARF-containing ring nucleases, polar interactions of a conserved Glu or Asp residue have been proposed to position their catalytic loop for activity (67). To further determine whether Glu122A side chain coordination of Arg98 main chain and/or negative charge of Glu122A side chain around central phosphoryl (P2b) prevents ring nuclease activity...
in Csa3Sso, we tested cA4 hydrolysis activity of the E122A and E122Q mutants, respectively. However, no ring nuclease activity was observed for these mutants (Fig. S5). Overall, our Csa3Sso data are consistent with the proposal that CARF motif 2 is involved in cA4 binding and not catalysis (67).

Solution structure reveals cA4-induced wHTH domain rearrangements in Csa3Sso

The high overall similarity of the Csa3Sso•cA4 complex structure with the apo Csa3Sso structure (61) suggested no significant Csa3Sso conformational changes upon cA4 binding (Fig. S2). This was perplexing to us specially since conformational differences in Csa3Sso could be the only other way to understand functional relevance of the cA4 binding given cA4 did not change the oligomerization state of Csa3Sso in our SV-AUC experiments. Also, cA4 binding has been previously shown to induce conformational changes in Can1 nuclease receptor of cA4 (69). To determine if the Csa3Sso•cA4 crystal structure accurately represents the solution state of this complex, we conducted small-angle X-ray scattering (SAXS) analysis of Csa3Sso in the presence and absence of cA4. In the presence of excess cA4 (sufficient to saturate all cA4-binding sites), the shape of dimeric Csa3Sso changed significantly. The data for Csa3Sso both in the presence and absence of cA4 displayed linearity in the classical Guinier analysis (Fig. S6), with an observed increase in the radius of gyration ($R_g$) in both the Guinier and inverse Fourier transform (GNOM) analyses (Fig. 7 and Table 2). Mass calculations from this data are consistent with in-solution dimers from both the states (Table 2). By Pr analysis, these differences coincide with increases in $R_g$ and $D_{max}$ and a redistribution of interatomic vectors to greater values (Fig. 7A). The numerical values derived from these analyses correlate very well with SAXS measurements made previously for apo Csa3Sso (61). This suggested that there is a significant conformational difference between the bound and apo states of Csa3Sso in solution.

Model-independent analyses including Guinier, Kratky, Porod-Debye, and mass calculations indicate that the cA4-induced conformation of the dimer is not due to changes in flexibility and disorder, or mass, but rather discrete differences in the configurations of the structural domains (Figs. 7B and S6 and Table 2). Although low in resolution, SAXS analysis allows for the rigorous testing of atomic models against their solution properties. These model-independent analyses would indicate that single atomistic models (ab initio or atomistic) could be reliably tested against the solution data to discern the nature of these conformational changes.

Crystallographic Csa3Sso models only provide atomic inventory for 212 of the 248 a.a. in the His$_6$-tagged Csa3Sso construct used in this study, including the disordered C terminus ($\sim$25 a.a.). To model this missing atomic inventory against its solution data, we employed the CORAL program, which uses coarse-grain beads for missing amino acids in a hybrid bead-atomistic modeling approach (70). When the method was applied, Csa3Sso in its unbound form could be readily reconciled against the SAXS data ($\chi_{\text{Crysol}} = 1.5$). No improvement was observed when allowing for the C-terminal wHTH domain positions to be refined via rigid-body docking, without symmetry constraints (Fig. 7C). In contrast, a significant preference was shown for a large asymmetric wHTH domain positioning over its symmetric crystallographic configuration in the cA4-bound state ($\chi_{\text{Crysol}} = 2.9$ versus 1.1) (Fig. 7D). Therefore, the binding of cA4 to the Csa3Sso dimer was found to induce large asymmetric conformational changes in the position of the two wHTH domains in solution.

Figure 6. Csa3Sso lacks ring nuclease activity. No changes in the time-dependent stability of cA4 were observed in the reaction mixture containing 5 μM Csa3Sso and 500 μM cA4 in buffer G (see Experimental procedures) at 60 °C for 3 h. Buffer G has previously been used for detecting the ring nuclease activity in Csm6 (84). The reactions were analyzed by C18 HPLC analysis after quenching and deproteination by phenol–chloroform extractions.
These in-solution conformational changes involved asymmetric, but significant, rotations and displacements of the wHTH domains from chains A (wHTHA) and B (wHTHB) (Fig. 8). More specifically, the wHTHA domain exhibited an 83.1° rotation and 20.9 Å displacement (left inset in Fig. 8), whereas the wHTHB domain exhibited a 107.6° rotation and 29.8 Å displacement (right inset in Fig. 8). This results in what we refer to as a “closed” Csa3Sso state. We envisage that Csa3Sso samples between this closed and open (equivalent to apo conformation) states. Although it is possible that apo Csa3Sso populates a complex population of open and closed states, the closed conformation is stabilized upon cA4 binding. Overall, our data demonstrate significant allosteric repositioning of wHTHA and wHTHB domains upon cA4 binding, which may underlie cA4 regulation of Csa3 signaling.

**cA4 does not affect specific DNA binding by Csa3Sso**

Csa3asis (a Csa3Sso ortholog from *S. islandicus* REY15A) has been shown by Liu *et al.* to regulate transcription by binding to...
c4A-induced allostery in Csa3

Table 1
Structural data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data</th>
<th>Csa3Sso–c4A complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein Data Bank ID</td>
<td>6WXQ</td>
</tr>
<tr>
<td>Data collection</td>
<td>P2222</td>
</tr>
<tr>
<td>Space group</td>
<td>74.79, 118.83, 64.05</td>
</tr>
<tr>
<td>α = β = γ (Deg)</td>
<td>90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.92009</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>35.67 (2.09–2.05)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>9.2 (37.1)</td>
</tr>
<tr>
<td>Id/I(0)</td>
<td>19.6 (5.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (97.9)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>8.0 (7.4)</td>
</tr>
<tr>
<td>Total number of observed reflections</td>
<td>69,259 (3464)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>36,695 (1810)</td>
</tr>
<tr>
<td>CCl1/2 (%)</td>
<td>99.8 (92.3)</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>53.9</td>
</tr>
</tbody>
</table>

Refinement

| Rwork/Rfree (%)                           | 16.0/20.52          |
| Bond length (Å)                           | 0.008               |
| Bond angle (Deg)                          | 0.998               |
| Number of atoms                           | 3371                |
| Protein                                   | 284                 |
| Ligand                                    | 112                 |
| Water                                      | 284                 |
| Ramachandran plot                         | 99.03               |
| Additionally allowed (%)                  | 0.97                |
| Outlier (%)                               | 0                   |

Rmerge = Σ |Fo| − |Fc|/Σ |Fo|, calculated with a working set of reflections. Rfree is Rwork calculated with only the test set with 10% of reflections. Data for the highest resolution shell are given in parentheses. The structure was determined using single crystals.

a pseudopalindromic consensus region upstream of CRISPR leader and spacer acquisition operon (viz., cas1, cas2, cas4, and cas1/cas3a genes) (59, 60). We therefore evaluated the effect of the c4A-induced repositioning of whHTHs observed in our c4A-bound solution structure on the DNA binding by Csa3Sso. Our superpositional docking structure of the DNA fragments from homologous OhrR–DNA complexes (PDB ID: 1Z9C) (71) onto these apo state structures showed no change in DNA conformations for both wHTHA and wHTHB. More specifically, the fragment docked onto the whHTHA in the closed state sterically clashed with the CARFb, and the one docked onto whHTHB showed an ~90° rotation with respect to its position in the apo state model (Fig. 9A right panel). Since each docked DNA fragment here represents one of the two palindromes Csa3Sso is expected to bind, we considered experimental evaluation of a possibility of DNA binding

Table 2
Parameters derived from SAXS analysis of Csa3Sso

<table>
<thead>
<tr>
<th>Sample</th>
<th>Concentration (mg/ml)</th>
<th>qRg</th>
<th>Rg (Å)</th>
<th>Dmax (Å)</th>
<th>P*</th>
<th>MM (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Csa3Sso</td>
<td>8.3</td>
<td>0.38–1.23</td>
<td>28.7 ± 0.1</td>
<td>100</td>
<td>3.9</td>
<td>58.9 (56)</td>
</tr>
<tr>
<td>Csa3Sso + 180 μM c4A</td>
<td>6.7</td>
<td>0.23–1.28</td>
<td>28.4 ± 0.1</td>
<td>101</td>
<td>3.9</td>
<td>58.9 (56)</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>0.40–1.26</td>
<td>30.8 ± 0.2</td>
<td>110</td>
<td>4.0</td>
<td>66.1 (56)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.44–1.34</td>
<td>31.3 ± 0.2</td>
<td>110</td>
<td>4.0</td>
<td>66.4 (56)</td>
</tr>
</tbody>
</table>

Errors reported reflect the uncertainty in the value for Rg determined using classical Guinier fitting.

To further analyze whether binding of c4A affected specific DNA binding by Csa3Sso, we performed MST-based binding affinity analysis of Csa3Sso using PCas4a as a ligand where we removed precipitated material before the fluorescence measurements. This analysis depicted a Kd of 1.62 ± 0.17 μM for binding of Csa3Sso to PCas4a, which is moderately better than that observed previously for Csa3Sso binding to an analogous S. islandicus cas4a promoter (57). However, the addition of excess of c4A (or cA6) did not significantly change the PCas4a binding affinity (Kds of 3.20 ± 0.36 and 2.4 ± 0.17 μM in the presence of cA4 or cA6, respectively) (Fig. 9D).

In conclusion, these analyses showed that binding of c4A increases solubility of the Csa3Sso in the presence of DNA. However, in our assays using the selected binding regions, the specific DNA binding affinity of Csa3Sso was unaltered in vitro

J. Biol. Chem. (2022) 298(2) 101591
by adding cA4 alone. Although we might have missed the physiologically relevant DNA target of Csa3Sso, this also brings up possibilities of alternate models of transcriptional regulation discussed below.

Discussion

The type I interference complexes are highly efficient in clearing infections associated with phage genomes containing an intact PAM sequence (72). The phages, on the other hand, have evolved escape strategies like the generation of escape mutations in the PAM sequences (1, 48, 73, 74) and production of anti-CRISPR (Acr) proteins (75, 76). During infection by a resilient phage, a coexistent type III interference (csm/cmrr) complex works independent of the type I Acr- and PAM motif to facilitate phage clearance (Fig. 1) (25, 48, 49, 77, 78). However, excessive activity of the csm/cmrr complexes and cOA receptors results in nonspecific genomic mutagenesis and RNA hydrolysis, respectively (54, 79). The type III interference therefore needs to prevent this aberrant fate during reinfection by the same virus. Consistent with this, the type III interference is bioinformatically predicted to trigger de novo spacer acquisition and crRNA production enabling the type I system for a future reinfection (Fig. 1, A and B) (4, 25). However, the regulatory mechanisms underlying such a reversal to type I interference are not clear.

A CARF domain is a variant of the Rossmann fold lacking the canonical G-X-G-X-(G/A) motif involved in binding NAD(P)H or FADH2 (80). Instead, CARF domains generally conserve a (D/N)-X-(S/T)-X2-(R/K) motif in their βN4–αN4 loop that is known to bind cOAs in many nucleic acid hydrolases (81). The βN4–αN4 loop is conserved in the Csa3 transcription factors (residues marked with white stars in red background in Fig. 5), which, along with the βN1–αN1 and βN5–βN6 loops, creates a potential nucleotide-binding pocket (61). However, biochemical and structural characterization of ligand binding specificity and regulation of Csa3 family transcriptional factors has been missing (4, 57, 61). Despite the low micromolar binding affinity we observed for cA4 binding to Csa3, we expect it to be physiologically relevant since high micromolar cA4 concentrations have been estimated to be attained in an infected Sulfolobus cell. More specifically, every phage transcript molecule detected by the type III interference complex produces an intracellular cA4 concentration of ~6 μM. Thus, a concomitant synthesis of multiple phage RNA molecules could result in an intracellular cA4 concentration in the multiples of six to the number of RNA molecules detected (64). Therefore, a relevant biological scenario includes failure of type I CRISPR systems leading to the concomitant detection of several phage transcripts by the type III interference complexes, which then produce cA4 at a rate that significantly exceeds its hydrolysis by the ring nuclease effectors. As a low-affinity cA4 receptor, Csa3Sso may therefore act only as a last resort to regulate spacer acquisition, crRNA synthesis, and DNA repair. Such a mechanism could help alleviate cellular toxicity known to increase upon extensive spacer acquisition (37). This is further supported by the fact that Csa3Sso lacks any demonstrable ring nuclease activity in vitro (discussed below), which could reduce the effectiveness of such a system.

The previously predicted nucleotide-binding pocket in Csa3Sso involved 12 CARF domain residues. Seven of these constitute the motif 1 (T7–Φ9–G9–F10–(D/N)11–E12–X4–R17), which, along with E122, forms two symmetry-related pocket walls, and the other four comprise motif 2 (L93–X2–G86–Φ97–R98), which serves as the floor of this pocket (Φ stands for a hydrophobic residue) (61). Of the 12 residues from these regions, eight were found at the Csa3Sso:cA4 structural interface, whereas four (namely, Thr7, Glu12, Arg17, and Leu93) were present outside of the interface (Figs. 4 and 5). Furthermore, we identified six additional Csa3Sso residues at this interface (Thr13, Phe14, Pro35, Val39, Thr42, and Gly123). Of these, Phe14 is significantly conserved among Csa3 homologs (Fig. 5) and, along with Phe10 from motif 1, contributes substantially to hydrophobic stacking of cA4 adenines (Fig. 4C). Similar π–π
stacking contributions of a Trp residue (Trp42) in a *Thermus thermophilus* Can1 receptor of cA4 have previously been proposed to stabilize cA4 in an asymmetrical conformation (69), which may also be true for the extended cA4 conformation bound to Csa3 Sso. Furthermore, the essential interactions of the highly conserved Csa3 Sso-Gly96 with the terminal cA4 phosphoryl moieties may additionally contribute to this stabilization (Fig. 4). Finally, our observed contribution of Csa3 Sso-Arg98 to the central phosphoryl moieties in cA4 is consistent with a previously reported loss of the binding of the cA4 analog to a multisite Csa3bSis mutant encompassing an Csa3 Sso-Arg98-equivalent residue (57). These residues' functional relevance in Csa3a and Csa3b homologs for binding cA4 further confirms the validity of our

Figure 9. cA4 does not significantly affect specific DNA binding by Csa3 Sso in vitro. A, predictive docking of DNA fragments (orange backbone) onto the wHTH domains of SAXS-derived apo (blue protein model on left) and cA4-bound (light brown protein model on right) Csa3 Sso dimers. The DNA fragment docked onto protomer A in the cA4-bound Csa3 Sso dimer sterically clashed with a region in the CARF domain from protomer B in this dimer (encircled by a red dotted circle). The DNA fragment docked onto protomer B in this cA4-bound dimer exhibited a ~90° rotation instead. To obtain these models, wHTH domains of *B. subtilis* OhrR (PDB code: 1Z9C (71), RMSD: 1.7 Å) were individually superimposed onto wHTH domains of the Csa3 Sso dimer and OhrR was omitted from the view for clarity. A single and straight model of B-form DNA with two copies of the palindromes could not be modeled with the apo Csa3 Sso dimer owing to a slight misorientation (not shown) between Csa3 Sso wHTH domains (possibly requiring a bend in the DNA). B, precipitation of Csa3 Sso in the presence of PCas4a DNA and its solubilization by cA4 binding to the Csa3 Sso CARF domain. Before taking the photograph, 50 μM Csa3 Sso (WT or R98A mutant), 10 μM PCas4a, and/or 500 μM cA4 were added in the binding buffer without Tween 20 (see Experimental procedures) and incubated for 30 min at 20 °C. C and D, cA4 does not significantly affect the binding of Csa3 Sso to the PCas4a promoter in vitro. EMSA shows no effect of cA4 on the specific binding of Csa3 Sso to PCas4a (C) that drives spacer acquisition in *S. solfataricus* P2. Cy5-labeled PCas4a, 50 nM, was added to different Csa3 Sso concentrations ranging from 500 nM to 25 μM in the presence or absence of cA4 added in a 50:1 ratio of cA4:Csa3 Sso. D, microscale thermophoresis binding affinity analysis also shows no significant effect of cA4 (as well as cA6) on the Csa3 Sso binding to the PCas4a. For microscale thermophoresis, Csa3 Sso was held constant at 500 nM, and PCas4a concentration was varied from 763 nM to 25 μM in the absence or presence of 50-fold excess of cOAs (25 μM) to Csa3 Sso.
cA4-induced allostery in Csa3

observed Csa3Sso•cA4 structural interface to both these homologs. Therefore, based on our structure–function analysis, we propose following revised motifs facilitating cA4 binding to Csa3 proteins: ΦF5–GΠ–F10–(N/D)11–X1–C13–F14 (cA4 binding motif 1) and ΦF5–(R/I)9–G96–ΦF7–R98 (cA4 binding motif 2), where subscripted numbers denote residue positions in Csa3Sso, and the symbols ζ, π, and X denote hydrophilic, small, and noninteracting amino acids, respectively (Fig. 5) (82). The cA4 binding motif 2 interacts primarily with the phosphoryl groups, whereas the cA4 binding motif 1 engages all the adenine rings in cA4 (Fig. 4, A and B). Of interest, the cA4 binding motif 2 is also conserved in other CARF proteins including Csx1, Csm6, and SiRe_0244 (a ring nuclease) (61, 68, 83, 84), whereas the cA4 binding motif 1 is unique to the Csa3 homologs (Fig. 5).

A conserved Glu122 from the βN5–βN6 loop (outside of the above two motifs) was found to significantly limit the cA4 binding affinity in Csa3Sso, via electrostatic repulsion to the central cA4 phosphoryls; substituting Glu122 with an Ala, or more conservatively with a Gln residue, drastically increased cA4 binding affinity of Csa3Sso (~145- or 125-fold, respectively) (Figs. 4 and 5). The conservation of Glu122 during evolution further indicates that it might be limiting Csa3 response only to high cellular cA4 concentrations during type III phage interference. This is also supported by the lack of a self-limiting ring nuclease activity in Csa3Sso. Nevertheless, this represents an interesting example of a rational surface design that increases ligand binding affinity and has potential applications in engineering an amplified Csa3 response to phage infections.

Owing to variation in the identity of the catalytic residues in the active sites of the known ring nucleases, the mechanistic details of ring nuclease activity are still emerging (67). Except for Crn3 that requires metal ions for ring nuclease activity, most of the cOA receptors employ a metal-independent nucleophilic substitution mechanism where a general base deprotonates a ribosyl 2’ hydroxyl (attacking group) for a nucleophilic attack on the scissile phosphorus atom and/or stabilizes a pentacovalent transition state by coordinating scissile phosphoryl oxygens (68, 85). Ultimately, these interactions position the 2’ hydroxyl oxygen (O’), phosphorous (P), and 5’-ribose oxygen (O”) in-line (with an ideal O’-P-O” angle ~180°) for the phosphodiester bond hydrolysis (86, 87). Enterococcus italicus cA6 ring nuclease (EcSm6), however, is an exception where alternate residues in the protein sterically force cA6 O’-P-O” in a compatible in-line conformation (88). Although originally postulated to be involved in catalysis, the motif 2 Arg/Lys residues from CARF-containing ring nucleases have recently been found only to mediate cOA binding (67, 85). Instead, residues in another conserved motif GxS/T have been recently implicated in the catalysis, where Ser/Thr and a Trp residue conserved adjacent to this motif participate in the catalysis (68, 83–85, 88). Furthermore, a Glu or Asp residue conserved in a few CARF domain proteins coordinates the Gly to position the GxS/T motif adjacent to the scissile phosphoryl group, which is hypothesized to facilitate hydrolysis (67). Despite the essential interactions of the central cA4 phosphoryl (P2b) with the motif 2 residues in Csa3Sso (Gly96 and Arg98), it lacks a demonstrable cA4 ring nuclease activity (Fig. 6). This could be due, at least in part, the absence of the GxS/T motif in Csa3 proteins. Also, the active site interactions of Arg98 and Glu122 with the target cA4 phosphoryl (P2b) results in an O’-P’-O” angle of 94° presenting with a stereochemistry conducive for nucleophilic attack by the 2’ ribosyl hydroxyl (Fig. 4C). By contrast, the nonplanar (P2a) and distal phosphates (P1a and P1b) that have O’-P’-O” angles of 158°, 125°, and 141°, respectively, lack coordination by side chain of a basic residue to stabilize the transition state. The lack of ring nuclease activity in the Csa3Sso proteins, unlike most self-limiting cOA ring nucleases, could allow for a long-term potentiation of the cA4 signal in an infected cell.

It is intriguing why X-ray crystallography data did not depict cA4-induced conformational changes revealed by our SAXS analysis. This could be explained, at least in part, by immobilization of the wHTHA by its symmetry-related crystallographic interactions with CARF A&B from the next Csa3Sso dimer in our Csa3 Sso complex crystals. A significantly large interface constituted by these packing interactions could have driven selection of biologically infrequent Csa3Sso conformer we observed in the crystal structure (Fig. S9).

Csa3bSis (a Csa3b homolog) regulation of type I interference (cas) genes and CRISPR spacer acquisition complex has been recently studied (57, 58). In the absence of an MGE, the subtype I-A interference (cas) genes are kept repressed by Csa3bSis-mediated recruitment of the Cascade–crRNA complex at the PCas promoter. During MGE invasion, recognition of a protospacer sequence by the Cascade–crRNA complex facilitates its release from the PCas, resulting in derepression of the cas gene expression (58). Furthermore, ~10-fold increase in Csa3bSis binding affinity to the PCas1 promoter has been reported in the presence of the cA4 analog in vitro, which suggests a further repression of the adaptation gene expression upon phage transcript clearance by the type III interference complex (57). cA4-induced conformational changes we observed using SAXS bring DNA binding face of the wHTHA (specifically helix cC3 that is expected to dock into the major groove of DNA) in proximity to the CARF domain (αN3 and βN2) from chain B (Fig. 8 left panel). Although this could potentially disrupt Csa3Sso interactions with at least one copy of the two palindromes in its binding site (Fig. 9 right panel), our attempts to analyze the effects of cA4 on the binding of Csa3Sso to PCas4a promoter showed no significant change in specific DNA binding by Csa3Sso. Although this unexpected observation does not fully align with previously published transcriptional regulation of the acquisition operon in S. islandicus by Csa3bSis, it could simply be because the two Csa3a orthologs utilize alternate regulatory mechanisms. For example, cA4 binding could improve stability of DNA-bound Csa3Sso or alter its DNA binding mode in vitro, which is supported by our increased Csa3Sso–DNA complex solubility in vitro. Alternatively, Csa3Sso may need to interact with other
similar ligand(s) or novel protein partner(s) for its transcriptional regulation. The latter mechanism, however, would be more intricate and distinct from a more common and straightforward regulatory model where ligand binding alone regulates DNA binding affinity of transcription factors to recruit RNA polymerase. Such an alternate mechanism could further harness different binding partners to differentially regulate a vast variety of CRISPR loci that S. solfataricus genome encodes in comparison with S. islandicus. Need for a Csa3sso partner is also indicated by in vitro instability of the Csa3sso–DNA complexes we observed in the absence of cA4 in vitro (Figs. 9, B and C and S8C). Consistent with this, Csa3a proteins have been hypothesized to recruit transcription factor B to a noncanonical TATA box coexistent with the Csa3a promoters for spacer acquisition in Sulfolobales (59). Furthermore, such a cooperative interaction of Csa3sso with protein partners may also further improve Csa3sso binding affinities with cA4 and/or target DNA. Although we anticipate involvement of our observed C4-induced Csa3sso conformations to underlie any sort of functional regulation, an additional possibility of DNA-dependent oligomerization cannot be ruled out with the existing data. Therefore, more work needs to be done to appreciate the binding of cA4 to Csa3sso.

Finally, although possibilities always exist for other high-affinity ligands for Csa3 (that could also better regulate DNA binding by Csa3 in vitro), it is unlikely for a Csa3 CARF domain dimer with four binding pockets aptly engaged in recognizing a four-nucleotide ligand to accommodate a completely different ligand structure. In this context, it is of high interest to characterize Csa3a homologs from organisms that are reported to lack Cas10 and CRISPR-Cas loci to identify such possible alternate Csa3a ligands (67). Nonetheless, our data and coevolution of csa3 genes with CRISPR loci in most prokaryotes support a cross talk between the type III and type I CRISPR systems in the form of cA4 binding to Csa3a homologs. This may underlie the regulation of spacer acquisition, crRNA gene expression, and (type I) Cascade III and type I CRISPR systems in the form of cA4 binding to cA4 structural interface/C15

Experimental procedures
Cloning, expression, and purification
The full-length N-terminally His6-tagged constructs for cloning wildtype csa3sso (accession number: Sso1445) and csa3sso (R98A) mutant were synthesized from Twist Bio, Inc, and cloned into the Ndel and EcoRI sites of a pBB75 vector using In-Fusion Cloning (Takara Bio USA) generating plasmids pBB75(His6-Csa3sso_WT) and pBB75(His6- Csa3sso_R98A), respectively. For the generation of other csa3sso mutants targeting the Csa3sso–CA4 structural interface (F10A, F14A, G96A, E122A and E122Q), the His6-tagged csa3sso (WT) was first cloned into Ndel and EcoRI sites of pET21b plasmid by PCR amplification from pBB75(His6- Csa3sso_WT) using primers pET21(N-His6-Csa3sso_WT)_F and pET21(N-His6-Csa3sso_WT)_R (Table S2). The plasmid thus generated, pET21(His6-Csa3sso_WT), was then subjected to Q5 site-directed mutagenesis (New England Biolabs) using manufacturer’s protocol and custom mutagenesis primers (identified with the prefix pET21 in Table S2).

For protein expression, the recombinant His6-tagged plasmids (pBB75 [His6-Csa3] and pET21b [His6-Csa3]) were transformed into BL21(DE3)-pLysS E. coli cells and selected on LB plates containing 0.03 mg/ml Kanamycin (for recombinant pBB75 plasmids) or 0.1 mg/ml Ampicillin (for recombinant pET21b plasmids). For the expression of the native and mutant Csa3sso proteins, a single E. coli transformant colony was grown in ZYP-5025 autoinduction media (89) containing 0.03 mg/ml Kanamycin (for recombinant pBB75 plasmids) or 0.1 mg/ml Ampicillin (for recombinant pET21b plasmids) at 22 °C for 44 to 48 h. Cells were harvested and resuspended in buffer A (20 mM Tris-HCl pH 8.0 and 400 mM NaCl) containing 0.1 mg/ml DNase I (Sigma-Aldrich), 5 mM β-mercaptoethanol, and 1% v/v protease inhibitor cocktail (Sigma-Aldrich). Cells were lysed using EmulsiFlex-C3 (Avestin, Inc), and the lysates of His6-Csa3 were preheated for 30 min at 65 °C, respectively, and centrifuged at 14,000 rpm for 40 min to remove the cell debris.

The supernatants containing all the wildtype and mutant His6-Csa3sso preparations were applied to a HisTrap Fast Flow column (Cytiva Life Sciences, Inc) pre-equilibrated in buffer A. Following a wash with five column volumes (CVs) of buffer A, Csa3sso was eluted with a three-step gradient of buffer B (20 mM Tris-HCl pH 8.0 and 50 mM NaCl) to buffer C (20 mM Tris-HCl pH 8.0, 50 mM NaCl, and 500 mM Imidazole): (i) 0% to 10% (v/v) buffer C in ten CVs, (ii) 10% to 40% (v/v) buffer C in ten CVs, and (iii) 40% to 100% (v/v) buffer C in 20 CVs. The fractions containing Csa3sso were pooled and applied to a Mono-Q anion exchange chromatography column (Cytiva Life Sciences, Inc) equilibrated with buffer D (50 mM Tris-HCl pH 9.0). The flow-through containing Csa3sso was collected and subjected to a Cibacon Blue 3GA column (Sigma Aldrich) equilibrated with buffer D. Bound proteins were eluted using a linear gradient of buffer D and buffer E (50 mM Tris-HCl pH 9.0 and 1.5 M NaCl).

His6-tagged Csa3sso preparations obtained above were concentrated and applied onto a Superdex 200 16/70 gel filtration column pre-equilibrated with buffer F (20 mM Tris-HCl pH 8.0 and 50 mM NaCl). The Csa3sso-containing fractions were concentrated using an Amicon Ultra-10 kDa cutoff centrifugal filter (Millipore) and stored at ~80 °C.

Microscale thermophoresis
The ligand-binding specificity of Csa3sso (WT) and the role of Csa3sso residues in ligand binding were determined by microscale Thermophoresis (MST). Wildtype or mutant His6- Csa3sso (200 nM) was fluorescently labeled in 2× binding buffer (10 mM Na/K phosphate, pH 5.8, 10 mM MgCl2, 25 mM NaCl, and 0.05% Tween 20) using 100 nM RED-Tris NTA His-Tag labeling dye (Nanotemper Technologies, Inc) and incubated in the dark at room temperature for 30 min. Two millimolar stocks of synthetic ligands including cA3, cA4, and cA6 (Biolog Life Science Institute) or linear ribonucleotide
cA4-induced allostery in Csa3

5′-rCrArArArA-3′ (Bio-Synthesis Inc) were serially diluted in a 1:1 ratio with nuclease-free water. The labeled proteins were then mixed with ligands in a 1:1 ratio. Only cA4 was used in the MST experiments with the Csa3Sso mutants (2 mM stocks of cA4 were used for F10A, F14A, G96A, and R98A mutants, and 20 μM stocks were used for E122A and E122Q mutants). The mixtures of proteins and ligands were incubated in the dark at room temperature for 30 min. The precipitated material was removed by centrifugation at 6000 rpm for 5 min. The supernatants were then loaded into Monolith NT.115 Series Premium capillaries in triplicate, and the thermophoresis was detected with 40% excitation power and 40% IR-laser power for an on-time of 20 s at 25 °C.

The MST experiment and data analysis for ligand-induced changes in Csa3Sso binding affinity with P cas4a was performed as mentioned above except for the following: 1 μM of His6-Csa3Sso (WT) was fluorescently labeled in 2× binding buffer with 40 nM RED-Tris NTA His-Tag labeling dye in the absence or presence of cA4 or cA6. The labeling mixtures were incubated in the dark at room temperature for 30 min. A 50 μM stock of P cas4a was serially diluted in 1:1 ratios with nuclease-free water. The labeled proteins ± ligands were then mixed with diluted P cas4a in a 1:1 ratio and incubated in the dark at room temperature for 30 min. The supernatants obtained after removal of precipitated material were then loaded into Monolith NT.115 Series Premium capillaries in triplicate, and the thermophoresis was detected with 100% excitation power and 60% IR-laser power for an on-time of 20 s at 25 °C.

The binding affinities of oligonucleotides and P cas4a to Csa3Sso (WT and mutants) were analyzed according to the law of mass action in a standard fitting mode of MO.Afinity analysis software (version 2.3).

Crystallization, X-ray data collection, data processing, model building, and refinement

Our high-throughput crystallization screen using His6-Csa3Sso in the presence of 2-fold molar excess of cA4 identified a crystallization condition yielding cA4-dependent His6-Csa3Sso crystals. The crystallization condition was optimized to obtain crystals growing up to 400 μm size. Csa3Sso•cA4 complex crystals were produced by the vapor diffusion method at 20 °C using a 1:2 mixture of Csa3Sso•cA4 (300 μM His6-Csa3Sso and 600 μM cA4) in gel filtration buffer (buffer F: 20 mM Tris-HCl pH 8.0 and 50 mM NaCl) and well solution (0.1 M K2SO4, 0.1 M Na/K 5.8, 16% PEG3350). X-ray diffraction data were collected in NSLS-II using AMX beamline at wavelength 0.92009 Å. The diffraction data were indexed, integrated, and scaled in HKL2000 (90). The initial phase information was obtained by molecular replacement in phenix phaser (91) using the native Csa3 structure (PDB ID: 2WTE) as template. The crystal belongs to P2₁2₁2₁ and two molecules of Csa3 are in an asymmetric unit. A strong difference density was observed on inspection of the F Cuomo-FC map at the CARF domain dimeric interface and was identified as bound cA4. The molecular model for the ligand cA4 (CHEBI:142457) was obtained from the ChEBI EMBL database (92). The ligand restraints were generated in Phenix ReadySet, and cA4 has been manually modeled on the difference density map. Iterative rounds of manual model building in Coot and refinement in Phenix refinement generated the final model with Rwork = 16.0 and Rfree = 20.5 (93, 94). The stereochemical quality of the final structure was verified using Ramachandran plot, and 99.03% of the residues are found to have favorable conformation, whereas only 0.97% of residues have allowed conformation; no outlier was found. The structure is submitted to PDB with a PDB ID 6WXQ. The data processing and refinement statistics are reported in Table 1.

Ring nuclease activity assays

The nuclease activity assays of cA4 were performed in two different conditions. In the first condition (used for Figs. S5, A and B), 50 μM Csa3Sso and/or 250 μM cA4 was incubated (in a cA4:Csa3Sso molar ratio of 5:1) at 55 °C for 3 h either in (i) 1× binding buffer or (ii) the buffer G (20 mM Tris-HCl, pH 7.5, 50 mM KCl, and 50 mM NaCl) previously used to demonstrate the ring nuclease activity of Csm6 (84). The reaction mixtures were deproteinized by ultrafiltration with an Amicon Ultra 3 kDa cutoff centrifugal filter (Millipore). In the second condition (used for Figs. 6 and S5C), 5 μM Csa3Sso (wildtype or its E122A/E122Q mutants) and/or 500 μM cA4 was incubated (in a cA4:Csa3Sso molar ratio of 100:1) in the buffer G or 1× binding buffer at 60 °C for 3 h. The reactions were quenched and deproteinized by phenol–chloroform extractions (84). For both conditions, the products and controls were collected and analyzed with High Performance Liquid Chromatography system (UltiMate 3000, Thermo Scientific) equipped with the C18 column (4.6 × 100 mm, 5 μm particle size, Thermo Scientific) and C18 column, (15 cm × 4.6 mm, 3 μm particle size, Supelco), for the first and second conditions, respectively.

Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments were performed at 20 °C with an XL-A analytical ultracentrifuge (Beckman-Coulter) and a TiAn60 rotor with two-channel charcoal-filled Epon centerpiece and quartz windows. Data were collected with detection at 280 nm. Complete sedimentation velocity profiles were recorded every 30 s at 40,000 rpm. Data were fit using the c(S) implementation of the Lamm equation as implemented in the program SEDFIT (95) and corrected for S20,w. Direct fitting of association models was performed using SEDPHAT (96). Calculated hydrodynamic properties for homology models were determined using WinHYDROPRO (97). The partial specific volume (μ), solvent density (ρ), and viscosity (η) were derived from chemical composition by SEDNTERP (http://www.rasmmb.bbrl.org/). Figures were created using the program GUSSI (98). All measurements were performed in 10 mM Na/KPO4 (pH 5.8), 10 mM NaCl, 25 mM MgCl2.
Small-angle X-ray scattering data collection

SAXS data were collected at beamline 16-ID (LiX) of the National Synchrotron Light Source II (99). Data were collected at a wavelength of 1.0 Å in a three-collimation configuration, yielding accessible scattering angle with 0.013 < q < 3.0 Å⁻¹, where q is the momentum transfer, defined as q = 4π sin(θ)/λ, where λ is the X-ray wavelength and 2θ is the scattering angle. The data to q < 0.5 Å⁻¹ were used in subsequent analyses. Samples were loaded into a 1-mm capillary for ten 1-s X-ray exposures. All measurements were performed in 10 mM Na/KPO₄ (pH 5.8), 10 mM NaCl, 25 mM MgCl₂.

SAXS analysis

Data were analyzed in the program RAW (100). When fitting manually, the maximum diameter of the particle (Dmax) was incrementally adjusted in GNOM (101) to maximize the goodness-of-fit parameter, to minimize the discrepancy between the fit and the experimental data, and to optimize the visual qualities of the distribution profile.

Hybrid bead-atomistic modeling of Csa3 was performed using the program CORAL (70), where the known structure was fixed in composition and inventory not resolved by X-ray crystallography was modeled as coarse-grain beads. Ten independent calculations for each protein were performed and yielded comparable results. The final models were assessed using the program CRYSOL. The models were rendered using the program PYMOL (102).

Electromobility shift assays

Top and bottom DNA oligonucleotide strands used for EMSA were purchased from Sigma-Aldrich (Table S2). The oligonucleotide strands were annealed by mixing them in 1:1 molar ratio followed by heating at 98 °C for the unlabeled, or 70 °C for the Cy5-labeled, DNA fragments for 15 min. Slow cooling to room temperature was used to anneal the fragments. Unlabeled probe-containing samples used for Fig. S8C were electrophoresed using a 2% (w/v) agarose Tris-Borate-EDTA (TBE) gel at 100 V for 30 min at room temperature. Probes were stained with ethidium bromide and visualized using Gel Doc XR+ Molecular Imager (Bio-Rad). EMSAs shown in Figure 9C were performed using the Cy5-labeled P_Cas4a probe. A 5’ amino-modified P_Cas4a top strand was obtained from Biosearch Technologies. The Cy5-coupled oligo fraction was purified from unlabeled oligos and loose dye in C18 reverse-phase HPLC using already established protocols (103, 104). The Cy5-labeled probes were electrophoresed in a 5% acrylamide TBE gel at 100 V for 1 h at 4 °C. DNA bands were visualized using FluorChem R gel imager (Protein Simple, Inc).

Data availability

Coordinates and structure factors for the Csa3_a00_cA4 structure have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org) with the accession code 6WXQ. Strains and plasmids are described in this article, and the raw data for the binding analyses in Figures 1D, 4E and 9D and Fig. S1 are available upon request.

Supporting information—This article contains supporting information (59, 61, 101, 102).

Acknowledgments—Analytical ultracentrifugation analyses were performed at the Johnson Foundation Structural Biology and Biophysics Core at the Perelman School of Medicine (Philadelphia, PA) with the support of an NIH High-End Instrumentation Grant (S10-OD018483). The crystallographic data were obtained at the 17-ID-1 beamline, whereas the SEC-SAXS data were obtained at 16-ID (LIX) at the National Synchrotron Light Source II, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704.

Author contributions—V. P. conceptualization; P. X. and K. G. methodology; P. X., A. D., K. G., and V. P. validation; P. X., A. D., and K. G. formal analysis; V. P. data curation; V. P. writing – original draft; P. X., A. D., K. G., M. B., and V. P. writing – review & editing; M. B. and V. P. supervision; M. B. and V. P. project administration; V. P. funding acquisition.

Funding and additional information—This research was supported by National Institutes of Health Grant R35GM119504 (to V. P.). K. G. acknowledges the support of the Johnson Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: cA4, cyclic tetra-adenylate; CARF, CRISPR-associated Rossmann fold; Cascade, CRISPR-associated complex for antiviral defense; cOA, cyclic oligo-adenylate; Crn, CRISPR ring nuclease; Csa3, CRISPR Apenn 3; Csa3biso_sulfolobus islandicus Csa3b; Csa3_sao_Saccharolobus solfatarius Csa3; CV, column volume; EMSA, electromobility shift assay; MGE, mobile genetic element; MST, microscale thermophoresis; PAM, protospacer adjacent motif; RNP, ribonucleoprotein; SAXS, small-angle X-ray scattering; SV-AUC, sedimentation velocity analytical ultracentrifugation; wHTH, winged helix-turn-helix.

References

cA4-induced allostery in Csa3


