# ALLOSTERIC ACTIVATION OF THE CRISPR-ASSOCIATED TRANSCRIPTION FACTOR CSA3 BY CYCLIC TETRA-ADENYLATE ($cA_4$)

by

Alexander Anthony Charbonneau

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2020

# ACKNOWLEDGEMENTS

The accomplishment that this thesis signifies was made possible by the help and support I've received from my teachers, co-workers, and family. I owe many thanks to my advisor, Dr. Martin Lawrence, and to my committee members Dr. Brian Bother and Dr. Mark Young. Thank you to the other members of the Lawrence Lab – Colin Gavin, Nate Burman, and Sharidan Brown – for your help in achieving this. Finally, thank you to my family members for your support, especially my wife, Kasey Kane Charbonneau, who pushed me to complete this and supported my every step along the way.

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

The CRISPR-Cas immune system provides adaptive and heritable immunity to archaea and bacteria to combat viral infection, and is a source of biochemical tools to researchers. This work combines structural biology and biochemical approaches to provide insight into mechanisms prokaryotes use to control the CRISPR-Cas immune system, linking subsystems into a coordinated response.

The first structure of *S. solfataricus* Csa3 determined by Lintner et al. revealed a dimer with a C-terminal wHTH DNA-binding domain and an N-terminal CARF domain with a putative ligand binding site predicted to bind a two-fold symmetric molecule with both negatively charged and hydrophobic/aromatic moieties, such as dinucleoside polyphosphates or nucleic acid molecules.[1] Later, analysis by Topuzlu et al. of the *A. fulgidis* Csx3 structure containing a 4 base RNA molecule in a binding pocket revealed similarities between Csx3 and the Csa3 CARF domain, and suggested CARF proteins could bind cyclic or pseudosymmetric linear RNA tetranucleotides represented by the ring-shaped RNA density in the Csx3 binding pocket.[2] Functional studies with *S. islandicus* Csa3 identified that SiCsa3 regulates transcription of acquisition genes (cas1, cas2, and cas4) and several CRISPR loci.[3,4] Additionally, two groups simultaneously showed that the Type III surveillance complexes produce cyclic oligoadenylate messengers, including cyclic tetra-adenylate ($cA_4$), which allosterically regulate the RNase activity of Csm6 and Csx1, other CARF proteins.[5,6] These advances support the original predictions by Lintner et al. and suggest that Csa3 binds a cyclic RNA as proposed by Topuzlu et al. Binding of $cA_4$ likely allosterically causes conformation changes in the wHTH, and regulates the protein's transcriptional regulation.

We present the crystal structure of *S. solfataricus* Csa3 complexed with cyclic tetraadenylate ($cA_4$). $cA_4$, as predicted,[1] is bound in the CARF domain 2-fold symmetric pocket, which stimulates conformational changes in the C-terminal domain. Additionally, we identify the presence of a palindromic predicted binding motif upstream of the Type I-A(2) acquisition cassette and CRISPR loci C and D, and reveal through EMSA analysis that Csa3 binds dsDNA nonspecifically with high affinity. Finally, ring nuclease activity is not detected in Csa3, suggesting longer term potentiation of the $cA_4$ in Csa3 than observed for Csx1/Csm6.[5,6]

# INTRODUCTION

The constant struggle between virus and host results in a constant cycle of evolution as one side alters recognizable features to evade the other or develops mechanisms for infecting of combatting infection. Being the numerically dominant biological entity on earth,[7] viruses have the advantage of probability in infecting cells that could otherwise spatially evade infection, thus prokaryotes have developed complex defense mechanisms for identifying and destroying invading nucleic acids. Although the restriction-modification and DNA phosphorothioation host defense systems offer a level of protection, the CRISPR-Cas (clustered regularly interspaced short palindromic repeats) system provides host immunity that is both adaptive and heritable.[8,9]

## Overview of the CRISPR-Cas Immune System

The basic tenets of the prokaryotic CRISPR-Cas immune system have been well defined.[8,9] Found in most archaea and many bacteria, this immune system provides protection from foreign or invading nucleic acids through a three-stage response; 1) adaptation, 2) crRNA biogenesis and maturation, and 3) interference. During the first adaptation stage acquisition genes (Cas1, Cas2, and Cas4) integrate short DNA strands of invading nucleic acids (protospacers) into CRISPR arrays within the host genome, a process that makes this immune system both adaptive and heritable (Figure 1.1). The CRISPR array/locus contains these new spacers as well as others inherited from the parent cell, each of which is separated by short palindromic “repeats” giving rise to the namesake Clustered Regularly Interspaced Short Palindromic Repeats or CRISPRs.

Transcription of the CRISPR array into pre-CRISPR RNA initiates the second stage, after which the new transcript is cleaved into individual spacers by Cas6 or other cellular nucleases, yielding CRISPR-RNA (crRNA). In the third stage the crRNA and Cas proteins are assembled into surveillance complexes that search out invading nucleic acids using the crRNA as a guide to interrogate the nucleic acids for complimentary base pairing. When the complex identifies an invading nucleic acid, nuclease proteins that are part of the complex or recruited cellular nucleases cleave the target, destroying the invading genetic material.

Diversity of CRISPR-Cas Systems

The diversity and arrangement of the *cas* genes involved in the stages of an immune response serve as guides for the delineation of CRISPR-Cas systems which are divided into 2 classes (I and II) and then further divided into 6 types and then into a multitude of subtypes each. This divergence stems from the variation in interference complexes, but also to a lesser degree from type/subtype-specific accessory proteins whose functions are yet unidentified, but are predicted to serve a variety of purposes such as signaling and control of the immune system. The work here focuses on the Class 1 systems (Figure 1.2) encoded in the *Sulfolobus solfataricus* P2 genome, which include three type I-A, three type III-B, and one Type III-D clusters as well as seven CRISPR arrays designated CRISPR loci A-G (Figure 1.3).[1,10,11] Cascade complexes, the Type I surveillance complexes (or aCascade for Type I-A specifically[1]), interrogate double stranded DNA in a PAM-dependent manner.[9,12] Upon recognition of a complimentary sequence, a Cas3 nuclease is recruited to degrade the invading DNA. Type III interference complexes do

not require a PAM and target the mRNA of transcriptionally active DNA.[13] Target recognition results in degradation of the DNA within the transcription bubble by Cas10 and cleavage of the mRNA by Cas7.

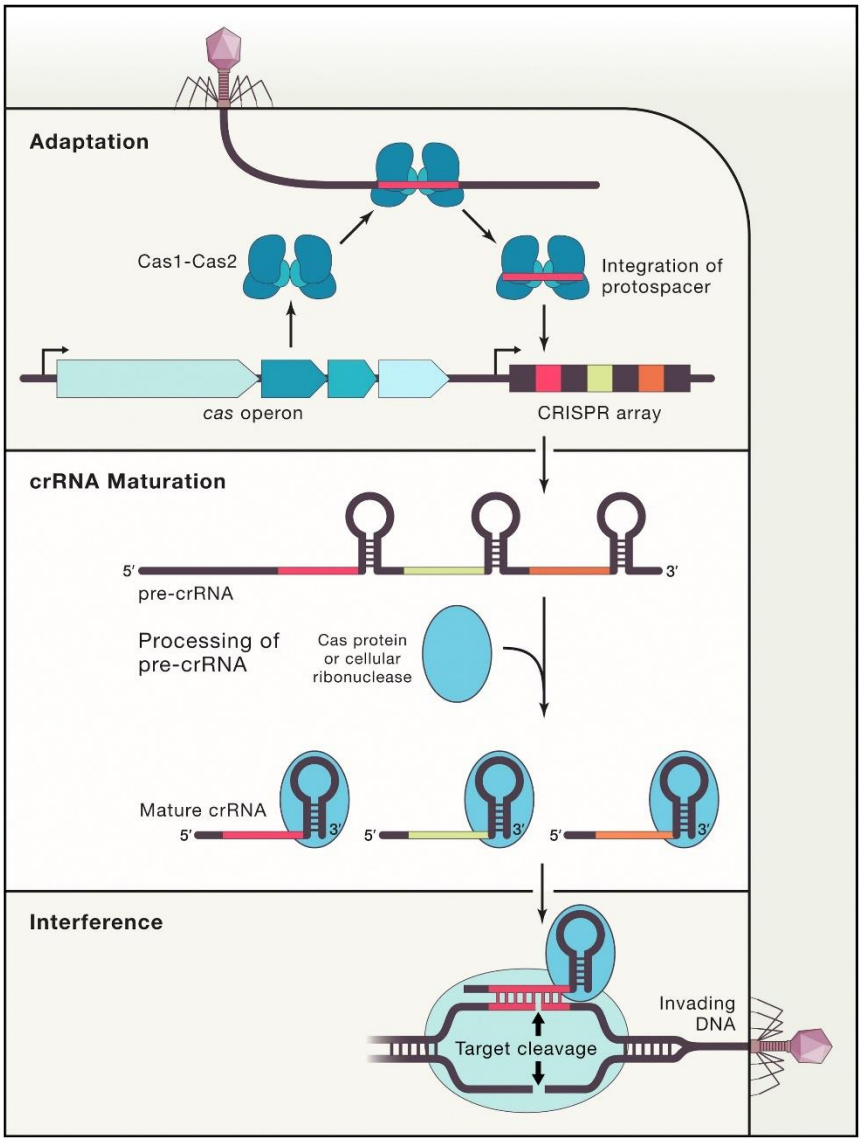


Figure 1.1 General overview of the stages of the CRISPR-Cas immune system. During adaptation, acquisition genes select a section of invading DNA and integrate it into the host CRISPR array/locus. The CRISPR array/locus is then transcribed into pre-crRNA, which is then cut at the repeat sequence into individual mature crRNAs by Cas6 or cellular nucleases. Each crRNA is then loaded into surveillance complexes that use the

crRNA to identify invading nucleic acids and degrade them. Figure used with permission from reference [8].[8]

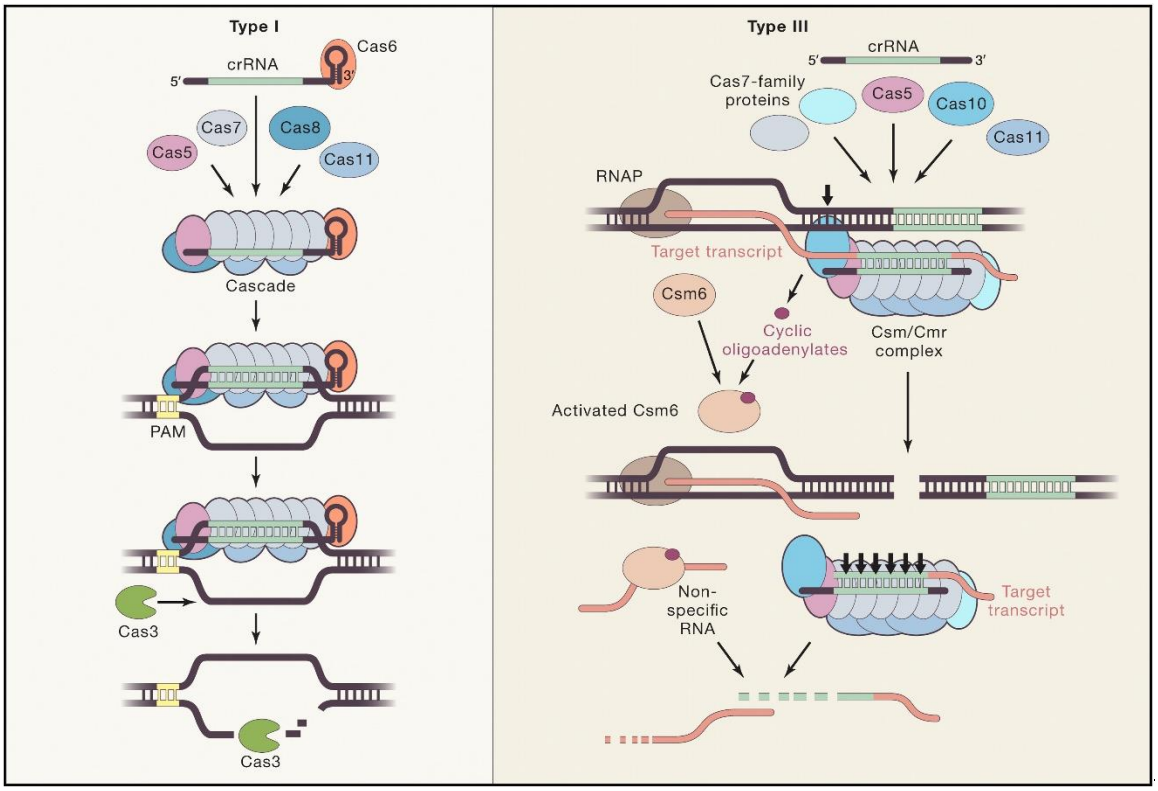


Figure 1.2. Type I and Type III Interference Complexes and Pathways. Both the Class 1 CRISPR-Cas systems utilize a complex of proteins in their surveillance complexes termed Cascade in Type I and Csm/Cmr complex in Type III. Cascade surveys dsDNA for a PAM, upon finding one, interrogates the DNA for complimentary base pairing with its crRNA, and recruits Cas3 to degrade the target[12]. In a PAM-independent manner, the Csm/Cmr complex surveys the mRNA of actively transcribed DNA for complimentary base-pairing to its crRNA and cleaves both nucleic acids upon identifying a target. Figure used with permission from reference [8].[8]

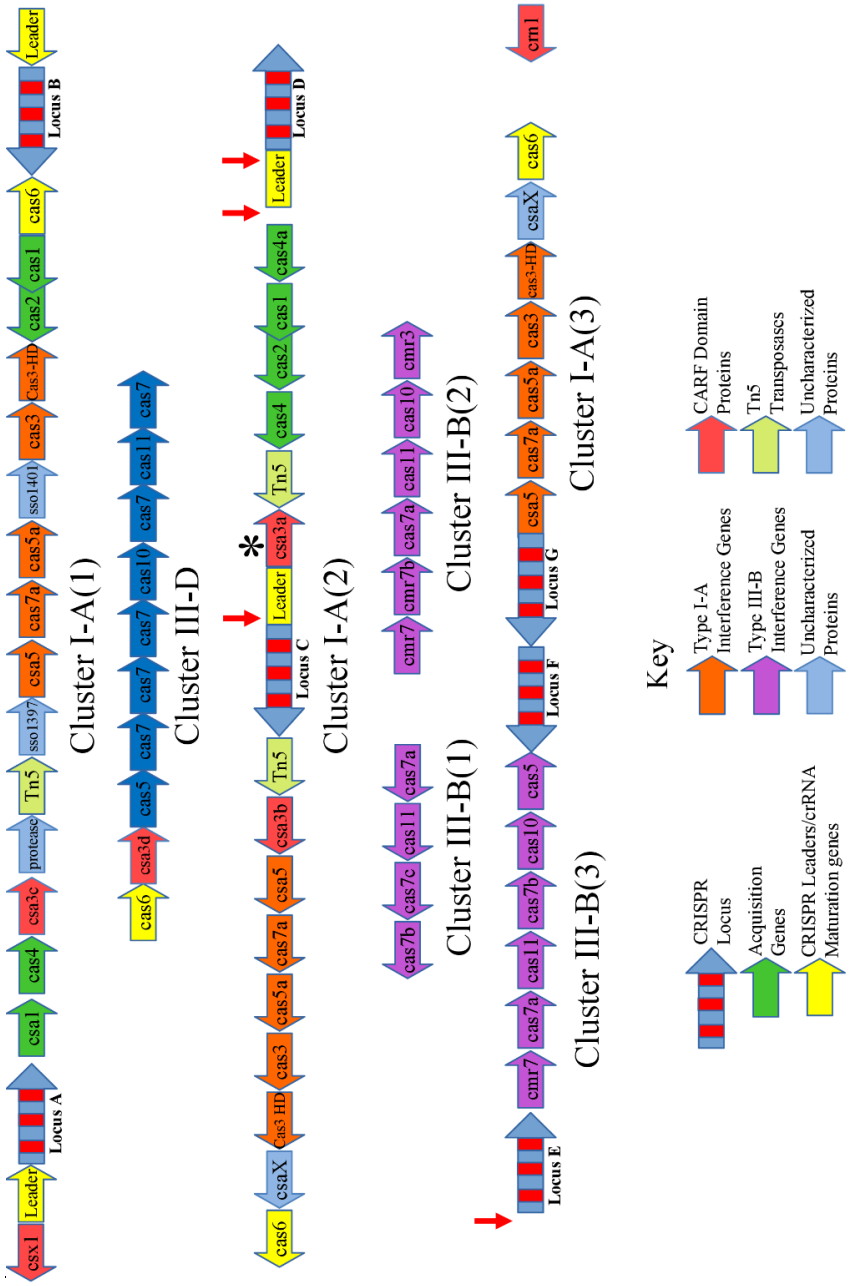


Figure 1.3. Arrangement of CRISPR-Cas Systems in the *S. solfataricus* P2 Genome. The genome contains seven CRISPR loci (blue/red striped arrows), three Type I-A cas clusters, three Type III-B cas clusters, and a single Type III-D cluster. Acquisition genes are represented in green, interference genes in orange/purple/Blue (Type I, Type III-B, and Type III-D respectively), CRISPR leader regions and crRNA maturation genes in yellow, uncharacterized genes in blue, TN5 transposases in light yellow, and CARF domain proteins in red. The location of Csa3a that this research focuses on is indicated by an asterisk. Locations of the predicted binding motif detailed in Chapter 4 are designated with a small, red arrow.

## CARF Proteins in Type I and III Systems

Ancillary proteins that have not been show to directly interact with interference complexes in type I-A, type III-A, type III-B systems include Csa3, Csm6 and Csx3 respectively.[14] As predicted by early bioinformatics work, the structure of Csa3 by Lintner et al.. revealed a two domain protein in which the N-terminal domain is a modified dinucleotide binding domain that continues into a C-terminal domain that forms a MarR-like winged helix-turn-helix (wHTH) domain.[1] The N-terminal domains of the dimer form a putative 2-fold symmetric binding pocket that spans the dimer interface opposite the C-terminal domains. Two conserved sequence motifs were identified in the putative N-terminal binding pocket: [Thr-h-Gly-Phe-(Asn/Asp)-Glu-X4-Arg] and [Leu-X2-Gly-h-Arg][1]. This suggested that the N-terminal domain may act as a ligand binding site that allosterically regulates the DNA-binding activity of the MarR-like C-terminal domains.[1] The putative ligand binding pocket contains conserved positively-charged residues (Arg98) and solvent-accessible aromatic residues (Phe10), suggesting the specific ligand that binds here may be a cyclic nucleic acid that bind symmetrically across the symmetric pocket due to the negatively charged phosphate groups and aromatic base moieties.[1] Analysis of the structure of another ancillary Cas protein Csx1 by Lintner et al. identified a similar fold and quaternary structure to the N-terminal domain of Csa3, suggesting the fold may be a conserved domain within CRISPR-Cas systems that regulates the activity of an attached C-terminal domain.[1,14,15] Further bioinformatic analysis grouped these shared N-terminal domains under the name CRISPR-Associated Rossman Fold (CARF) domain.[16] Additional structural, function,

and bioinformatics endeavors identified the C-terminal domains of Csx1 and Csm6 as metal independent HEPN ribonucleases that could be controlled by their CARF domains.[15–18] The presence of CARF domains in proteins containing varied N-terminal domains suggested that there would be a shared regulatory signal that activates the RNase activity of Csx1/Csm6, which could induce cell dormancy and/or cell death, and controlled transcription of other cas genes through Csa3.[14,16] The structure of another ancillary protein in *A. fulgidis*, Csx3, identified a cocrystallized 4-base RNA fragment in a binding pocket opposite from the metal-dependent exoribonuclease site.[19] Further analysis of the structure placed Csx3 in the CARF domain superfamily, which suggested that the common ligand for allosteric control of the other CARF proteins may be a cyclic 4-base RNA or a linear 4 base RNA bound pseudosymmetrically.[2,19]

Earlier studies on the Cas10 subunit of Type III surveillance complexes identified a cyclase domain whose function in the complex was unknown[20] until two separate groups in 2017 determined its role in producing cyclic oligoadenylates.[5,6] They found that the Cas10 cyclase domain synthesizes cyclic oligoadenylates from ATP, including cyclic tetraadenylate ($cA_4$) and cyclic hexaadenylate ($cA_6$), when the Type III surveillance complex identifies a target through complimentary base-pairing between the crRNA and the transcript mRNA. In addition to identifying a source of cyclic oligonucleotides, they showed that these products of the Type III surveillance complex stimulate the endonuclease activities of Csx1 and Csm6,[5,6] confirming the hypothesis that the common regulatory signal for CARF proteins is a cyclic oligonucleotide.[1,2,16] Interestingly, a level of complexity was added as different lengths of cyclic oligonucleotide were required to

stimulate the activity Csm6 in different organisms, *S. thermophilus* Csm6 requiring $cA_6$ and *T. thermophilus* and *E. italicus* Csm6 requiring $cA_4$ for activation.[5,6] A mutation in the CARF binding pocket of EiCsm6 abrogated the $cA_6$-stimulated RNase activity, confirming that the CARF domain was indeed responsible for the allosteric control.[5]

Following these discoveries, additional proteins in the CARF domain superfamily were identified that utilize the domain to bind and degrade $cA_4$, leading to deactivation of Csm6/Csx1 type III RNases.[21] These "Ring Nucleases" include Sso2081 (Crn1) and Sso1393.[21] The structure of Sso1393 (PDBid: 3QYF) reveals an N-terminal CARF domain and C-terminal wHTH domain[22], suggesting it is another isoform of Csa3 along with Csa3a and Csa3b, as such we are labeling it Csa3c as shown in Figure 1.3. The presence of ring nuclease activity in Csa3c along with the discovery of low level ring nuclease activity in Csx1 suggests that this autoregulatory catalytic activity may be present in many or all CARF domain proteins.[23]

The Role of Csa3 in Type I and Type III Systems

The role of Csa3 in *S. islandicus* Rey15A (SiCsa3) was uncovered with overexpression of the protein leading to increased transcript levels for cas1, cas2 and csa1, members of the acquisition cassette whose indirect overexpression resulted in enhanced acquisition of new spacers.[3,4] DNA foot-printing experiments identified a DNA binding site for SiCsa3 immediately upstream of the acquisition cassette containing cas1, cas2, and csa1 as well as in front of several CRISPR loci.[3] The locations of these binding sites suggests that SiCsa3 activates transcription of these acquisition genes to promote spacer acquisition and additionally promotes transcription of the CRISPR loci to

produce crRNAs from the newly acquired spacers.[3,4] The specific binding site and function for Csa3 in *S. solfataricus*, however, remains unknown.

## Research Goals

The work presented here is a component of an ongoing project to understand how Csa3 regulates the *S. solfataricus* CRISPR-Cas immune systems and how the Type I-A and Type III systems in the archaea interact to form a cohesive, global immune response. The project initially began with the determination of the Csa3 structure by Dr. Nathaniel Lintner,[1] and here this work characterizes the structure of the Csa3-$cA_4$ complex and the effect of $cA_4$ on DNA recognition. Additionally, this work elucidates the structural basis for $cA_4$ recognition by the CARF domain of Csa3 and defines the accompanying conformation changes that are predicted to activate transcription of the nearby acquisition cassette (cas4a, cas1, cas2, and cas4a) and CRISPR loci. Lastly, the identification of X-ray density for the complete ring structure of $cA_4$ suggests the lack of ring nuclease activity in Csa3 in our crystal form, meaning Csa3 may stay active longer during the signaling process than other CARF domain proteins with similar affinities to $cA_4$.

# MATERIALS AND METHODS

## Large Scale Expression and Purification of Csa3

The Gateway® Cloning system (Invitrogen) was used to prepare *S. solfataricus* Csa3 for expression in *E. coli*. A pDonor201 vector containing the Sso1445 gene that encodes Csa3 with an N-terminal 6x-His tag(pEntr201-6xHis-Sso1445), was obtained from laboratory stocks made by Nathaniel Lintner.[1] Using a site-specific recombination LR reaction, the gene was transferred into a pDEST14 vector, creating the pEXP14-6×His-Sso1445 expression vector, which was sequence verified (Nevada Genomics).

Csa3 was expressed and purified as previously described by Lintner et al..[1] Briefly, BL21(DE3)-pRIL *E. coli* (Stratagene) were transformed with the pEXP14-6×His-Sso1445 expression plasmid[1] and a single colony was used to inoculate 5 ml of LB media with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol and grown overnight. 750 ml of ZYP-5052 autoinduction media[24] containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol was then inoculated with 750 μL of the starter culture and grown at 37 °C for 16–20 h. Cells were harvested by centrifugation at 5500 x g (Sorvall Superspeed RC-2) for 20 min, and the pellets were stored at −80 °C.

Cell pellets were thawed and resuspended in 5 ml of lysis buffer (20 mM Tris pH 8.0, 400 mM NaCl, 0.1 mM PMSF) per gram of cell pellet. Cells were lysed by passage through a French press (American Instrument Co., Inc., Silver Springs, MD). The lysate was incubated at 60 °C for 20 min to denature *E. coli* proteins and clarified by centrifugation at 30,000 x g for 30 min. The supernatant was then applied to a gravity-flow column containing a 0.5- to 1-ml bed volume of Ni-NTA agarose (Qiagen). The

column was washed with 8 column volumes of wash buffer (20 mM Tris, 400 mM NaCl, and 10 mM imidazole, pH 8.0), and Csa3 was eluted in 10 mM Tris (pH 8.0), 50 mM NaCl and 300 mM imidazole. Csa3 was then applied to a calibrated Superdex 75 (GE Healthcare) column equilibrated with 10 mM Tris (pH 8.0) and 50 mM NaCl and fractions containing the major peak were combined. Protein concentrations were determined by Bradford assay[25] using Protein Assay Reagent (Bio-Rad) and bovine serum albumin as a standard. The purity and molecular weight of Csa3 were confirmed by SDS-PAGE (Figure 2.1). Csa3 without measurable contaminants was purified with a yield of ~3 mg pure protein per gram of cell pellet after separation on the Superdex 75 column. Stocks of Csa3 in with 10 mM Tris (pH 8.0) and 50 mM NaCl were stored at -80°C at a concentration of 1 mg/mL, which yielded ~70% soluble protein after thawing at 37°C. It is unknow if Csa3 can be stored in the Ni-NTA elution buffer.

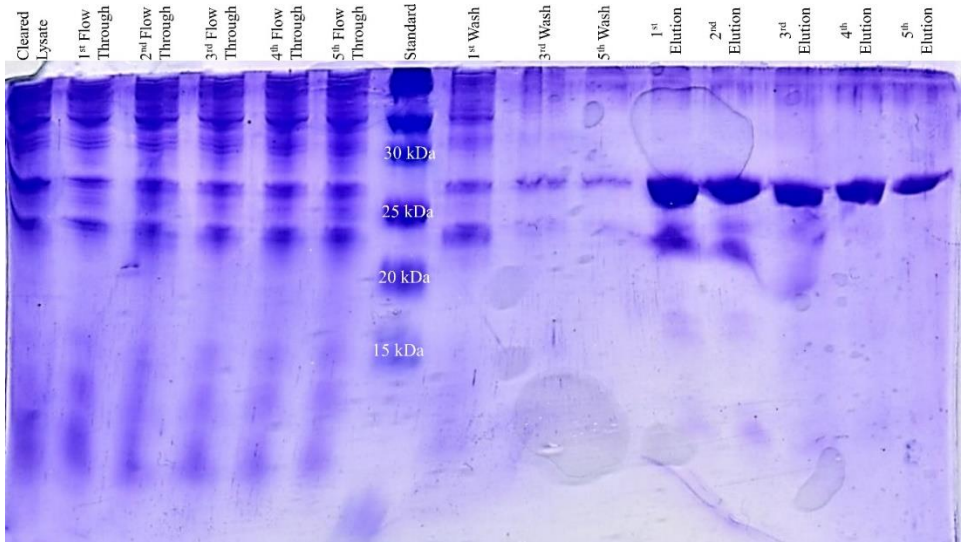


Figure 2.1. SDS-PAGE Gel of Fractions from the Purification of Csa3. 15% Acrylamide Gel. Csa3 runs at ~27 kDa in elution fractions with a ~18 kDa contaminant that disappears after the first 2-3 elution fractions. All contaminants are removed after separation on a Superdex 75 column.

## Screening for Crystallization Conditions

To attempt to co-crystallize Csa3 in complex with $cA_4$, purified Csa3 was concentrated to 15 mg/mL with 10,000 MWCO Amicon Ultra™ spin concentrators (Millipore) for the initial conditions tested – those used by Lintner et al.[1] to crystallize the Csa3 in the absence of $cA_4$ (2 uL 15 mg/mL Csa3 and 2uL 37.5-42.5% MPD, 5% PEG-8000, and 0.1 M sodium cacodylate at pH 7.0-7.5) with an additional 0.2 uL 10 mM $cA_4$ (c-tetraAMP, BioLog Lifescience Institute) in hanging drop vapor diffusion at 22°C.[1] Addition of $cA_4$ under these conditions caused immediate precipitation of protein.

We then attempted to soak $cA_4$ and $A_4$>P into crystals of the protein grown under the conditions used by Lintner et al..[1] 0.2 uL of 10 mM $cA_4$ or $A_4$>P (ChemGenes Corporation) was added to drops containing fully grown crystals in hopes that the ligand would bind to the crystallized protein. Addition of either RNA resulted in rapid crystal decay, yielding no usable crystals for data collection.

After neither initial co-crystallization nor ligand soaks succeeded in producing usable crystals, we screened a variety of new co-crystallization conditions. We reduced the protein concentration to 7.5 mg/mL, at which point immediate precipitation did not occur upon adding RNA. These new screens were set up using a TTPLabtech Mosquito® Crystal crystallization robot. Conditions were screened by sitting drop vapor diffusion consisting of 100 nL Csa3 at 7.5 mg/mL, 50 nL 2mM $cA_4$, and 200 nL of well solution from a variety of crystallization screens. Crystallization screens included Hampton Crystal Screen 1, Hampton Crystal Screen 2, Hampton PEG Ion 1, Hampton PEG Ion 2, Hampton Salt Rx 1, Hampton Salt Rx 2, Hampton Crystal Screen Lite,

Hampton Crystal Screen Cryo, Hampton Peg Rx 1, Hampton Peg Rx2, Hampton Index, and Emerald BioStructures Cryo. Control screens were set up using the same conditions for the protein and well solution but replacing the $cA_4$ with 50 nL water. Drops were incubated at 22 °C for 1-2 days and multiple conditions yielded crystals and microcrystals, though few were successfully scaled up to 4 uL hanging drop vapor diffusion to generate larger crystals for data collection.

## Final Crystallization Conditions

The condition that yielded the final crystals used for data collection were set up at 22 °C using 2 µl of SsCsa3 at 7.5 mg/mL, 2 µl of well solution (14–22% PEG MME 550, 100 mM imidazole pH 7.0, and 150 mM malate at pH 7.2), and 0.2 uL of 10 mM $cA_4$. Co-crystals of Csa3 with $cA_4$ up to 0.30 mm × 0.075 mm × 0.075 mm in size were obtained in 3 to 5 days (Figure 2.2).

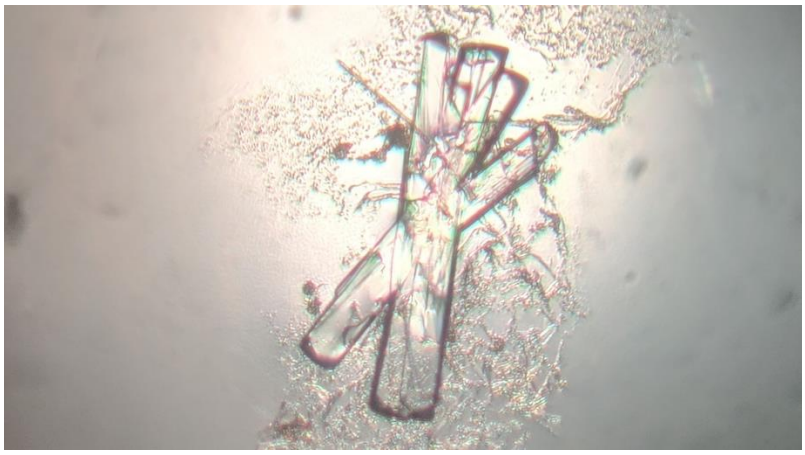

Figure 2.2. Co-Crystals of *S. solfataricus* Csa3 and $cA_4$. These crystals were grown in 18% PEG MME 550, 100 mM imidazole pH 7.0, and 150 mM malate at pH 7.2. Csa3 was present at 7.5 mg/mL and $cA_4$ at 1 mM. Single crystals were easily separated from the cluster with a cryo loop. The precipitate around the crystals forms on the surface of the drop, but was tangled with the crystals while looping out another from the drop.

X-Ray Data Collection

Single crystals of the Csa3-$cA_4$ complex were flash frozen in liquid nitrogen and 360° of data was collected at 160 K using our MicroMax-007 X-ray generator and Rigaku R-axis $4^{++}$ image plate detector. Csa3-$cA_4$ crystals diffracted to 2.45 Å resolution (Figure 3.1). Data were indexed, integrated and scaled in space group $P2_12_12_1$ (a = 46.1 Å, b = 93.4 Å, c = 105.2 Å) using the HKL2000 software package.[26] Data quality parameters are presented in Table 1.

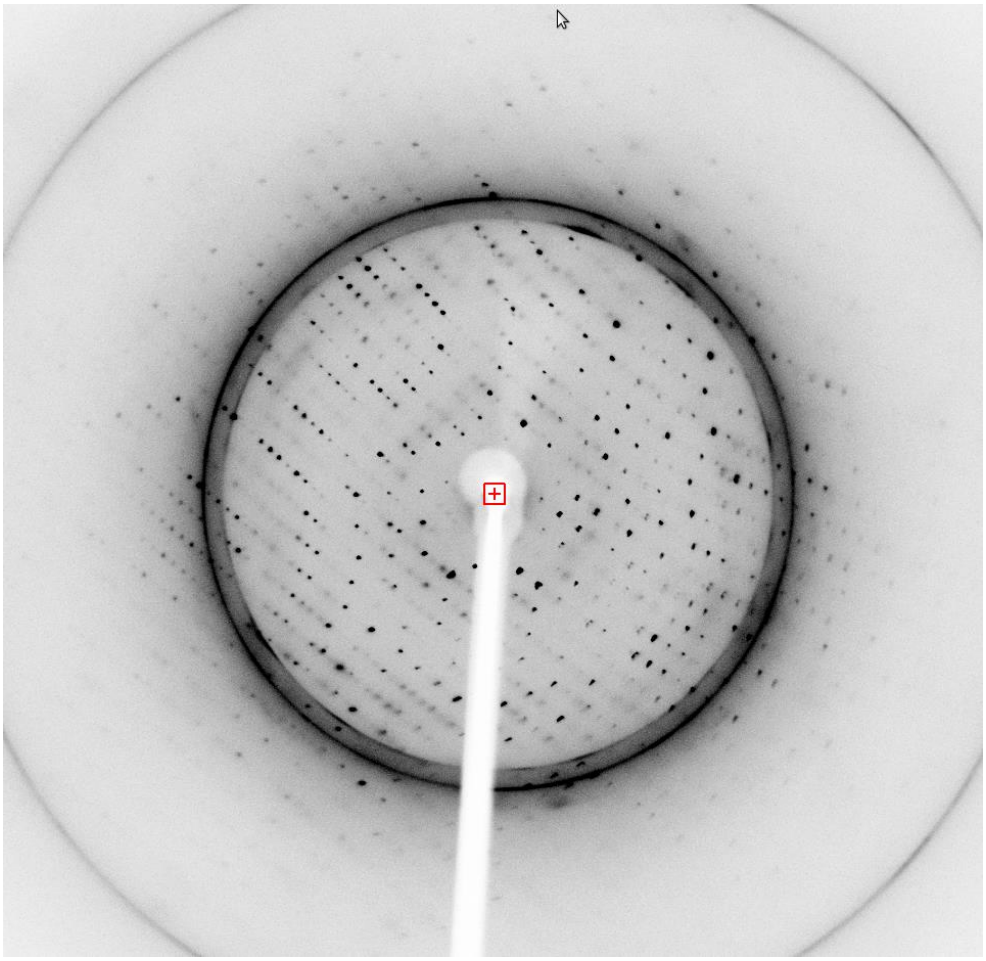

Figure 2.3. Diffraction image from the Csa3-$cA_4$ crystal used in the final data collection. This is the first image from the dataset used to solve the structure of Csa3 in complex with $cA_4$. The mouse arrow points out a reflection at 2.46 Å.

**Table 1**
**Data Collection**

| | |
|---|---|
| Data Set | SsCsa3-$cA_4$ |
| Wavelength (Å) | 1.54 |
| Space Group | $P2_12_12_1$ |
| Cell Constants (a,b,c; Å) | 46.1, 93.4, 105.2 |
| ($\alpha$, $\beta$, $\gamma$; $^o$) | 90, 90, 90 |
| Resolution Range[a] (Å) | 27.91-2.45 (2.49-2.45) |
| Unique Reflections[a] | 16,551 |
| Average Redundancy[a] | 10.4 (10.4) |
| $I/\sigma$[a] | 36.4 (2.5) |
| Completeness (%) | 96.8(100) |
| $R_{sym}$[a,b] (%) | 7.9 (82.0) |
| $R_{pim}$[a] (%) | 2.6 (26.5) |
| CC1/2[a] | 0.99 (0.737) |

[a]Numbers in parenthesis refer to the highest resolution shell.
[b]$R_{sym}=100*\Sigma_h\Sigma_i|I_i(h)-<I(h)>|/\ \Sigma_h I(h)$ where $I_i(h)$ is the $i^{th}$ measurement of reflection h and <I(h)> is the average value of the reflection intensity.

## Structure Determination and Refinement

The Phaser-MR module[27] of Phenix[28] was used to determine initial phases for Csa3 in complex with $cA_4$ by molecular replacement. The structure of Csa3 without the ligand (PDB ID: 2WTE)[1] was used as a search model. Due to expected conformational changes upon $cA_4$-binding, N- and C-terminal domains were placed separately. 2 copies of each domain were placed, yielding a single dimer within the asymmetric unit (40.5% solvent). The molecular replacement yielded a solution with a refined LLG of 2211, a TFZ score of 20.3, and an R-value of 42.4. The structure was completed by iterative model building with Coot[29] and refinement with phenix.refine.[28] Temperature/libration/screw parameters[30] were included in the refinement, with each of the two SsCsa3 chains and the single $cA_4$ ligand divided into twelve

temperature/libration/screw groups (chain A: 1–110, 111–157, 158–177, 178–112; chain B: 1–28, 29–55, 56–84, 85–97, 98–127, 128–157, 158–185, 186–212; $cA_4$: 1–4). The final $R_{work}$ and $R_{free}$ were 20.3% and 24.3%, respectively. MolProbity was used for model validation,[31] with 96% of residues falling in the most favored regions of the Ramachandran plot and none in disallowed regions. The overall MolProbity score[31] was 1.51, placing SsCsa3- $cA_4$ in the 99$^{th}$ percentile for overall geometric quality among protein crystal structures of similar resolution (2.458 ± 0.25 Å). Residue numbers in the model are consistent with the native, nontagged protein sequence. Residues 160–162, 174–176, 192–195, and 213–237 of chain A, and 213-237 of Chain B were not modeled due to lack of interpretable electron density. Additional details on model refinement and model quality are presented in Table 2. The model has been deposited in the Protein Data Bank (6W11). Structural figures were generated with Pymol.[32]

| **Table 2**<br>**Model Refinement** | |
|---|---|
| $R_{work}$[c] (%) | 20.3 (28.8) |
| $R_{free}$[c] (%) | 24.3 (34.6) |
| Real Space CC[d] (%) | 84.8 |
| Mean B Value (overall; Å$^2$) | 85.0 |
| Coordinate Error (based on maximum likelihood, Å) | 0.37 |
| RMSD from ideality: | |
| Bonds (Å) | 0.005 |
| Angles (°) | 0.687 |
| Ramachandran Plot[e] (%) | |
| Most Favored (%) | 96.0 |
| Additional Allowed (%) | 4.0 |
| PDB Accession Code | |
| [c]$R_{work} = \Sigma\|\|F_o\|-\|F_c\|\|/\Sigma\|F_o\|$ where $F_o$ and $F_c$ are the observed and calculated structure factor amplitudes used in refinement. $R_{free}$ is calculated as $R_{work}$, but using the "test" set of structure factor amplitudes that were withheld from refinement (10%). | |
| [d]Correlation coefficient (CC) is agreement between the model and $2mF_o$-$DF_c$ electron density map. | |
| [e]Calculated using MolProbity.[33] | |

Generation of EMSA Fragments

Based on the predictions found in the results section, we tested 8 sequences in electrophoretic mobility shift assays (EMSA): the primary 350 bp region in front of the cluster I-A(2) acquisition cassette (cas4/cas1/cas2/cas4a), 350 bp fragments of the leader

sequences for CRISPR loci C, D, and E that contain the motif, and 175 – 350 bp sections of the leader sequences for CRISPR loci B and F. EMSA fragments were generated by PCR from *S. solfataricus* P2 genomic DNA using the primers listed in Table 3. A 350 bp control fragment lacking the predicted binding motif, BRE, and TATA box was generated from the Lambda phage genome. PCR products were the purified by QIAquick PCR Purification Kit (Qiagen).

## Electrophoretic Mobility Shift Assays

To test the affinity of Csa3 for the PCR amplified fragments, 12 uL EMSA binding reactions containing 5.2 ng/uL, varying concentrations of Csa3a, and either 1 uM $cA_4$ or no $cA_4$ were incubated at 37°C for 30 min in TAE buffer. A Quick-Load® Purple 100 bp DNA Ladder from New England BioLabs and the incubated samples were loaded onto a 2% w/v agarose gel with 10 ug/mL ethidium bromide. The gel was buffered with TAE buffer and bands were separated at 100 V for 40 min. The fluorescent bands were visualized and imaged using an Alpha Innotech AlphaImager 2200 gel imager. To acquire quantitative binding data, gel images were analyzed using ImageJ image analysis software. The bands of DNA that did not shift were isolated as individual "lanes" on the gel image using the "Analyze>Select Lanes" tool. After plotting the lanes with the "Plot Lanes" tool, each peak was constrained to the same width by drawing 2 vertical parallel lines on either side of the most intense peak (o nM Csa3) from the top to the bottom of the whole plot. This forced a boundary on either side of the peaks that both ensured the same area was measured for each and also gave a reference point for the two points used to eliminate background signal. To remove the background, a line was drawn between

the two points where the vertical boundaries intersected the intensity plot, creating an enclosed space from the peak. The total intensity for the enclosed space was then recorded using the "Wand" tool (Figure 2.4). To calculate the percent of the DNA bound for each band the total intensity of the band was subtracted from the total intensity of the 0 nM Csa3 band and the resulting intensity was divided by the total intensity of the 0 nM Csa3 band. The calculations for percent DNA bound assume that the total intensity of the 0 nM band represents 100% of the DNA loaded into the well. Using GraphPad Prism, a non-linear curve was fit to the calculated percent bound-DNA using the "Specific Binding with Hill Slope" function: $Y=B_{max}*X^h/(K_d^h + X^h)$ where Y is the percent bound-DNA, X is the corresponding Csa3 concentration in nM, $B_{max}$ is the maximum specific binding in percent (constrained to 100%), $K_d$ is the concentration of Csa3 in nM needed to achieve half-maximum binding, and h is the Hill coefficient.

Table 3
PCR Primers

| Fragment Name | Annotation | Size (bp) | Genome Location | Annealing Temp. °C | Primers (5' to 3') |
|---|---|---|---|---|---|
| Cas4a (II) | Promoter region for Cas4a Cluster I-A(2) | 350 | 1305123-1305472 | 63 | ATAACGTCTAGGCTTTATCGGAGG |
| | | | | | CGTTTGATGTTAGTTTCTTAGGAGGTA |
| Cas4a (I) | Promoter region for Cas4a/Cas4 Cluster I-A(1) | 342 | 1240163-1240504 | 63 | CCGCAACTAACCCGTTCC |
| | | | | | TTTATCCTCTTTGAAAGCAACAGC |
| Cas1 | Promoter region for Cas1/Cas2 Cluster I-A(1) | 350 | 1253317-1253666 | 63 | GAATGGAGTTCTGTAGGGGCAA |
| | | | | | TAATCGAAATGTAAGTGTGTAAATCACAG |
| LB | Promoter region directly upstream of CRISPR Leader B | 470 | 1260396-1260865 | 63 | GATATTATACCGGTTGAGCCTGC |
| | | | | | CAGCCTTAGTGAAATTAGGAAATATTACG |
| LC/D | Promoter region directly upstream of CRISPR Leaders C/D | 339 | 1305422-1305760 | 63 | TGTCCCGTTTTTGTAAGTGGG |
| | | | | | TCGATAGAAAGATTTAAATATGCGGAATAC |
| LE | Promoter region directly upstream of CRISPR Leader E | 270 | 1743799-1744068 | 64 | GGGAAAGAGTTCCCCCGTA |
| | | | | | ACGTCGCTGCCAATTTTCC |
| LF | Promoter region directly upstream of CRISPR Leader F | 175 | 1811222-1811396 | 62 | CTGTTAGTAAGCCCCTTGCTAC |
| | | | | | AGAGAATACTTGTTACCTAACTGGT |
| Lambda | Control Fragment from Lambda phage genome | 350 | 24398-24747 | 62 | TAGGAATTGGTTAGCAAGTTACTACC |
| | | | | | TTGCTAACAGGTATCGTTTGGAG |

Table 3. PCR primers used to amplify EMSA fragments from the S. solfataricus P2 genome. The fragment name denotes the name used to reference the indicated fragment in experimental procedures and results. 'Cluster I-A(1)' and 'Cluster I-A(2)' in the annotation refer to the CRISPR cluster as described in Figure 1.3. The genome location indicates the location of the amplified fragment within the genome of *S. solfataricus* P2 for all fragments besides Lambda-350 which is in the Lambda phage genome. The top primer listed for each fragment represents the forward primer while the bottom one represents the reverse primer.

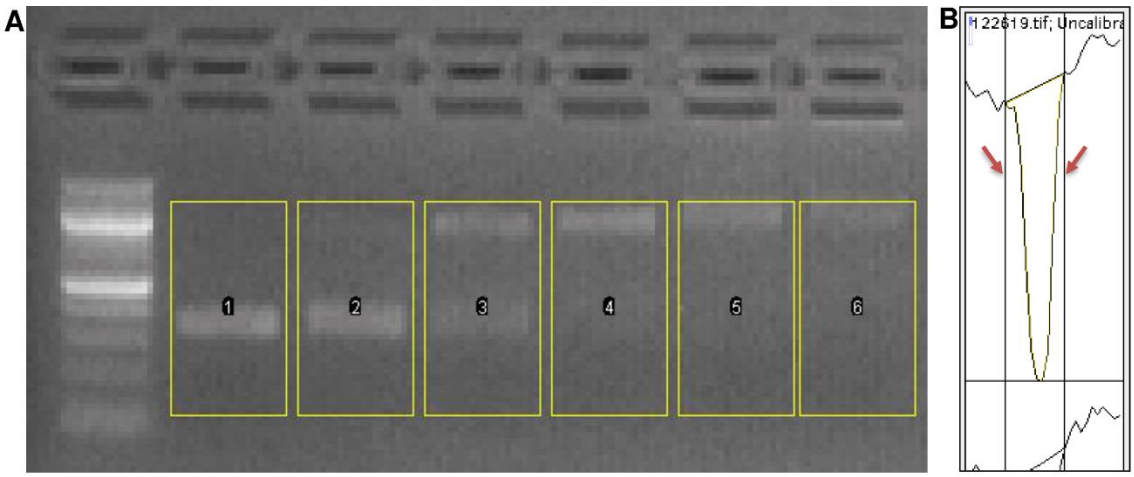


Figure 2.4. ImageJ Analysis of EMSA Gels. A. 32-bit image of the an EMSA gel with yellow boxes indicating the "lane" selection areas for analysis. Only the lower, unshifted band is necessary for quantitation. B. The "Plot Lanes" interface shows peaks for pixel intensity across the selected lanes (only lane 1 is fully visible in the figure). Red arrows indicate lines drawn perpendicular to the lane boundary in order to isolate the single peak that represents the unshifted band. The yellow-highlighted area indicates the area selected with the "wand" tool to calculate total intensity for the peak.

## Ring Nuclease Assay

Ring nuclease products were analyzed as described by Athukoralage et al. Briefly, 250 µM $cA_4$ and 25 µM protein were incubated at 40ºC for 7 hr in 10 mM Tris (pH 8.0) and 50 mM NaCl. The RNA was then isolated by phenol-chloroform extraction and 2 µL of the aqueous layer was spotted onto an Agela Technologies TLC Silica plate. The plate was placed in a sealed chamber containing 0.5cm of a mobile phase consisting of 70% EtOH, 30% $H_2O$, and 0.2 mM ammonium bicarbonate. The plate was developed until the solvent reached approximately 3/4 up the plate. The plate was then dried, and migrations were visualized under UV (254 nm).

# RESULTS

## The Co-Crystal Structure of Csa3 in Complex with $cA_4$

To elucidate the molecular basis for $cA_4$ recognition by Csa3 and the conformational changes induced by $cA_4$ binding, we determined the structure of Csa3 co-crystallized with $cA_4$. While the crystallization conditions were significantly different from those reported for the $cA_4$-free structure of Csa3 reported by Lintner et al, it crystallized in the same space group with similar unit cell parameters (Table 1). Consistent with this crystal form, the asymmetric unit contains a single Csa3 homodimer, and the overall structure is largely similar to the $cA_4$ free structure. However, the position of the wHTH domains relative to the CARF domains are noticeably different, and the conformational change appears to be driven by subtle conformational changes in the N-terminal CARF domain that are associated with $cA_4$ binding.

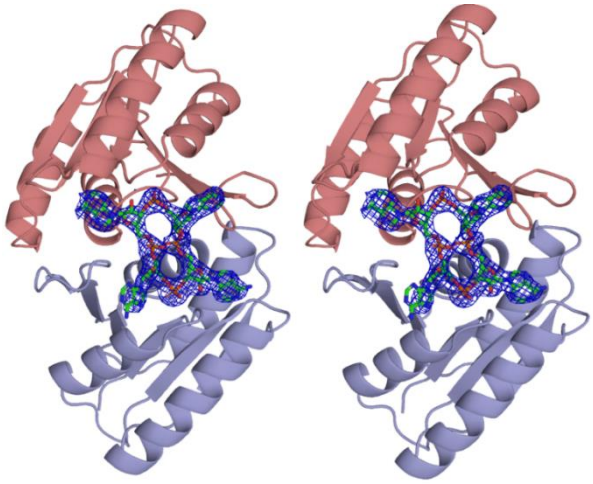

Figure 3.1. Omit map showing density for $cA_4$ in the Csa3 CARF domain. Ribbon diagram of the Csa3 dimer bound to $cA_4$ looking up into the $cA_4$ binding pocket. Chain A of the dimer is colored blue and chain A is colored red. $cA_4$ is shown in a stick representation with C atoms colored green, N atoms in blue, O atoms in red, and P atoms in orange.

CARF domain recognition of $cA_4$

The N-terminal CARF domain retains its overall fold, a doubly wound, mixed β-sheet with flanking α-helices. Similar to a classic dinucleotide binding domain, the first 5 β-strands ($\beta_1$-$\beta_5$) run parallel and are connected by helical, right-handed crossovers. However, the fold then diverges from the classic "Rossmann fold", as $\beta_5$ makes a reverse turn into $\beta_6$, which runs antiparallel to $\beta_5$, giving the sheet a $\beta_3(\uparrow)$-$\beta_2(\uparrow)$-$\beta_1(\uparrow)$-$\beta_4(\uparrow)$-$\beta_5(\uparrow)$-$\beta_6(\downarrow)$ topology. The reverse turn into $\beta_6$ also results in the loss of a covering alpha helix. Critically, CARF domain dimerization restores this structural element, placing $\alpha_4$ of the second subunit across the $\beta_5/\beta_6$ end of the β-sheet. The $\alpha_4$ helices are thus buried at the dimer interface, where they run parallel to each other. Overall, the non-canonical β5/β6 reverse turn and loss of the connecting α-helix is critical to dimer formation and are thus

As expected,[1] a single $cA_4$ molecule is bound within a 2-fold symmetric pocket spanning the subunit interface at the "base" of the Csa3 dimer (Figure 3.1). The floor (or ceiling, depending on orientation) of this pocket is formed by neighboring β4α4 loops from each of the 2 subunits. Principal residues in this loop are Gly96 and Arg98. The Arg98 side chains run antiparallel to each other as they reach across to the subunit interface and hydrogen bond to the main chain carbonyls of Gly96 in the adjacent subunit. The unsaturated Gly96 main chain amines and the Arg98 side chains then mediate alternating interactions with the phosphate groups of $cA_4$. These interactions explain the strong conservation of these residues in "motif 2" of Lintner et al.[1] defining features of the CARF domain fold.

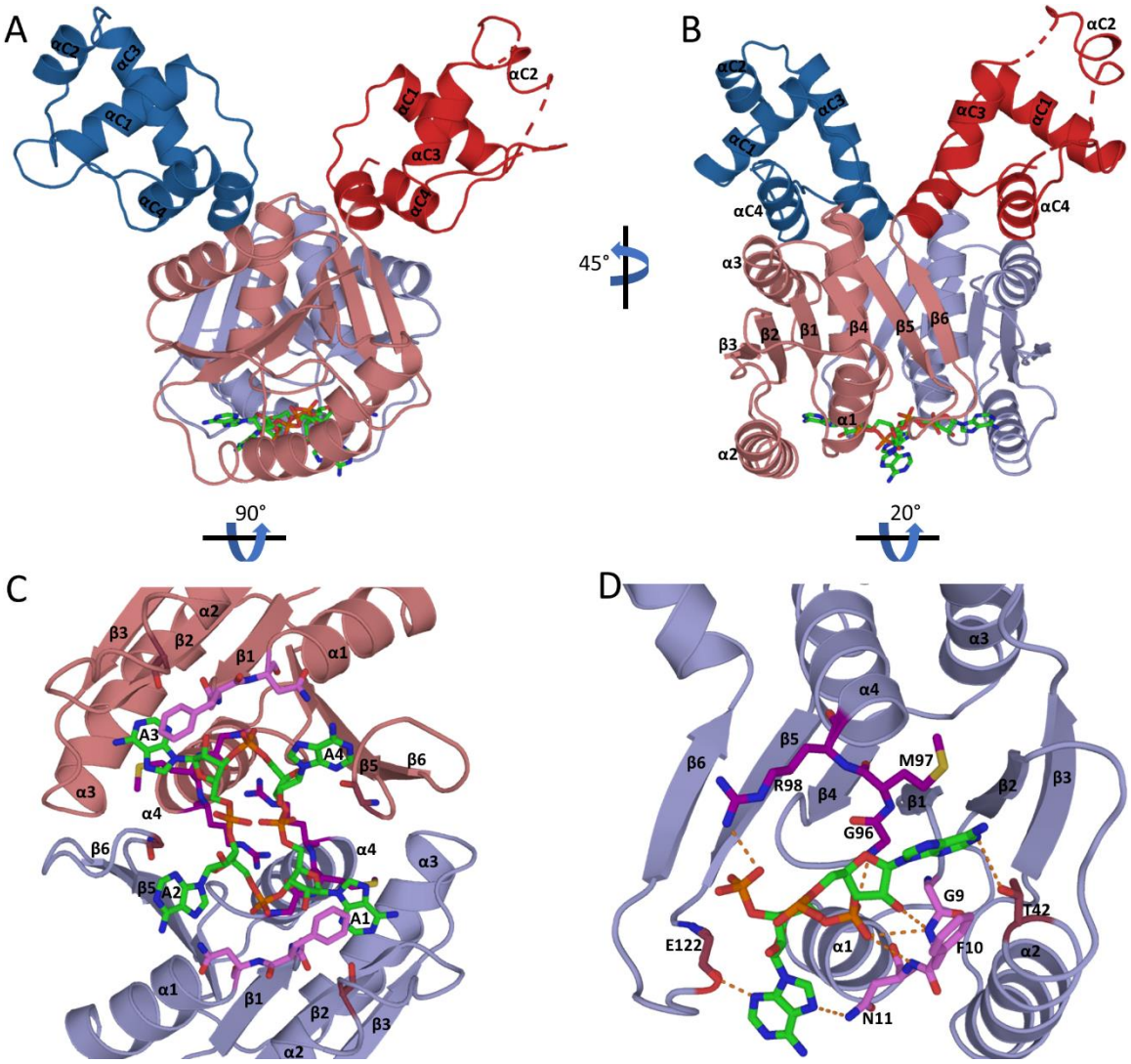


Figure 3.2. The Csa3 dimer binds $cA_4$ with a 2-fold symmetric binding pocket at the interface of the A chain in blue and the B chain in red. **A.** Ribbon diagram of the Csa3 dimer bound to $cA_4$. The N-terminal CARF domains that form the $cA_4$ binding pocket are colored light red/blue, the C-terminal wHTH domains are colored dark red/blue, and $cA_4$ is shown in a stick representation with C atoms colored green, N atoms in blue, O atoms in red, and P atoms in orange. **B.** The $cA_4$-bound Csa3 dimer rotated -45° about the Y-axis relative to the orientation in panel A shows the location of $cA_4$ in the binding pocket. Secondary structure elements are labeled on chain B (red). **C.** Ribbon diagram of the $cA_4$-bound Csa3 dimer looking into the $cA_4$ binding pocket. Relative to the orientation in panel A, the structure is rotated -90° about X. Motifs 1 and 2 of Lintner et al. (pink and purple respectively), T42 (deep red), and E122 (deep red) which make direct interactions with $cA_4$ are represented as sticks. **D.** Ribbon diagram of a single Csa3 subunit bound to bases 1-2 of $cA_4$. The structure is rotated -20° about X relative to B with the same representations as in C. Hydrogen bonds are shown as orange dashes.

$cA_4$ is bound in an oval-shaped conformation with all bases present in the anti conformation, radiating outward from the oval-shaped ribose-phosphate backbone. Further, each ribose is present in a 2'-endo confirmation. Consistent with this, Suitename[31] identifies the dihedral backbone conformations for nucleotides 2 and 4 at the distal ends of the oval as 6p conformers, and assigns nucleotides 1 and 3 as "outliers" of the 2o conformer.[34] Details for the RNA conformers are found in Table 4. The $\beta_1\alpha_1$ and $\beta_5\beta_6$ loops project above the floor of the $cA_4$ binding site in an alternating pattern about the dimer axis, where they successively interdigitate between each of the 4 adenosine moieties. Specifically, the $\beta_1\alpha_1$ loop inserts between bases 1 and 2 in subunit A, and between bases 3 and 4 in subunit B. Similarly, the $\beta_5\beta_6$ loop in subunit A rises between bases 2 and 3, and the same loop in the B subunit between bases 1 and 4 (Figure 3.2).

Table 4 - RNA Conformers for $cA_4$*

| Nucleotide | δ,δ,γ | Name | Suite | δ-1 | ε | ζ | α | β | γ | δ |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 22m | 2o | 2'emmtm2' | 158.08 | -145.90 | 50.51 | - | -119.33 | -71.42 | 156.24 |
| A2 | 22p | 6p | 2'epptp2' | 156.24 | -151.05 | 125.27 | 61.82 | 167.29 | 49.25 | 155.06 |
| A3 | 22m | 2o | 2'emmtm2' | 155.06 | -146.90 | 75.04 | 70.37 | 165.48 | -66.01 | 153.03 |
| A4 | 22p | 6p | 2'epptp2' | 153.03 | - | - | 74.01 | -172.64 | 55 | 158.08 |

*For definitions of the RNA backbone conformer nomenclature see Richardson et al.[34]

The first nucleoside is thus bound in a groove with walls formed by the $\beta_5\beta_6$ loop of subunit B and the $\beta_1\alpha_1$ loop of subunit A (Figure 3.2). The conserved Gly9-Phe10 in the $\beta_1\alpha_1$ loop [motif 1 of Lintner et al,[1]] play a pivotal role. The main chain NH of Phe10 hydrogen bonds to O2' of the ribose and the edge of the adenine base packs against the main chain of Gly9, allowing the Phe10 side chain to extend over the adenine ring. Val 39 of the $\beta_2\alpha_2$ loop also packs on top of the adenine ring, while the Met97 side chain ($\beta_4\alpha_4$ floor loop) provides a foundation underneath the adenine to complete a

hydrophobic binding pocket (Supplemental Figure 1). With the exception of Thr42 in the $\beta_2\alpha_2$ loop, which H-bonds to N1 of the adenine moiety, base specific interactions are not obvious.

In contrast to the first nucleoside, the second adenosine is bound entirely within subunit A, between the $\beta_1\alpha_1$ and $\beta_5\beta_6$ loops (Figure 3.2). The adenine ring settles on top of main chain atoms of the $\beta_5\beta_6$ loop (residues 122-124) and Phe14 of the $\beta_1\alpha_1$ loop, leaving the opposite face of the adenine solvent exposed. In addition, Thr13 accepts an H-bond from the adenine -$NH_2$ while the Asn11 side chain donates a hydrogen bond to N7 of the adenine ring. Finally, the Asn11 main chain NH hydrogen bonds to the 5'phosphate of this nucleotide.

Overall, then, the first subunit plays a predominant role in recognition of the first 2 nucleotides of $cA_4$. Similarly, the 2$^{nd}$ subunit largely mediates recognition of the last 2 nucleotides, using a similar set of interactions. However, the $cA_4$ is not bound in a completely symmetric fashion. Noticeably, a crystal contact with a neighboring asymmetric unit results in a cation-$\pi$ interaction between Arg200 and the solvent exposed adenine ring of the 4$^{th}$ nucleotide. This interaction appears to pull the adenine out of its binding pocket, breaking the interactions with Asn11 and Thr13. Further, the $\beta_5\beta_6$ loop near the crystal contact is also found in an alternate conformation relative to subunit A. Finally, each of the ribose moieties are present in a 2'-endo conformation and for residues 1, 2 and 3, the 2'-OH is hydrogen bonded to either O1P or O2P of the neighboring phosphate group. While the ribose in the 4$^{th}$ nucleotide is also 2'-endo, it reaches across the $cA_4$ molecule to instead hydrogen bond to the O2P of the 3$^{rd}$

nucleotide. This is possible, in part, because the phosphate group linking the 2nd and 3rd nucleotides does not sit down on the floor of the pocket, as it does for the 5'-phophate of the first nucleotide. Whether these differences are the result of crystallographic contacts, or there is inherent asymmetry in the solution bound form is not clear.

Conformational changes induced by $cA_4$

Local to the $cA_4$ binding site, the most prominent conformational changes are modest movement of the $\beta_1\alpha_1$ and $\beta_2\alpha_2$ loops as they move away from the dimer axis, expanding the conserved pocket to accommodate $cA_4$. At the same time, there are significant side chain rearrangements as Phe10 rotates over the adenine ring of the first (and third) nucleotide. These movements are accompanied by a small, yet significant rotation of the $\alpha_1\beta_2\alpha_2\beta_3\alpha_3$ side of the CARF domain (Supplemental Movie 2). Critically, the $\alpha_3$ helix runs antiparallel to the C-terminal helix of the winged helix-turn-helix DNA binding domain. This small rotation of the α3 helix is, in turn, transmitted to this C-terminal helix of the wHTH domain, acting as a pivot point for movement of the wHTH domain. Relative to the Csa3 dimer, the net result is a twisting movement of the two wHTH domains towards each other, into a conformation that is clearly more favorable for DNA binding, but would still require a slightly bent DNA conformation (Supplemental Movie 3). Noticeably, in the crystal, a neighboring Csa3 dimer packs between the two wHTH domains. Thus, crystal contacts may prevent the wHTH domains from adopting the most favorable DNA binding conformation.

Identification of A Palindromic DNA Motif

*S. solfataricus* P2 harbors 7 CRISPR loci (A-G) and six Cas clusters, of which three are Type I-A, three are Type III-B, and one is Type III-D (Figure 1.3). The clusters are designated here as I-A(1-3) and III-B(1-3), and are numbered according to their sequence in the genome. The Csa3 protein that this work focuses on is located in the I-A(2) cluster near the I-A(2) acquisition cassette (cas4a/cas1/cas2/cas4) and CRISPR loci C and D.

Working in a related organism, *Sulfolobus islandicus* strain Rey15A, Liu et al demonstrated that overexpression of Csa3 stimulates spacer acquisition.[3] In addition, chromatin immunoprecipitation and DNA foot printing studies identified potential DNA binding regions for Csa3 upstream of genes involved in spacer acquisition (csa1, cas2, cas2) and several CRISPR loci. Further, related sequences in other *Sulfolobales*, including *S. solfataricus* P2 were also reported.[3] Given the 2-fold symmetric nature of Csa3, we would expect a (pseudo-) 2-fold symmetric DNA binding site. We thus searched for a conserved palindromic motif within these suggested DNA binding regions in *S. solfataricus* P2. Indeed, a GGGC X10 GCCC palindromic motif upstream of cas4a and 3 of the CRISPR loci was identified by Colin Gauvin (Figure 4.1). Importantly, for the putative cas4a binding site, we identified a putative transcription factor B recognition element (BRE) and TATA box immediately upstream of the palindromic motif (Figure 4.1). This suggested Csa3 might act as a transcriptional activator, and that $cA_4$ might enhance the affinity of Csa3 for this putative csa4/cas1/cas2/cas4a promoter.

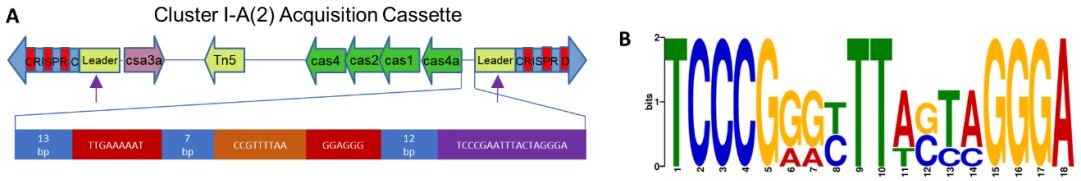


Figure 3.3. Composition of the cluster I-A(2) acquisition cassette. A. The cluster I-A(2) acquisition cassette lies between CRISPR loci C and D. Csa3a runs in the opposite direction as the acquisition genes (cas4a, cas1, cas2, and cas4), separated only by a predicted Tn5 transposase. Blue arrows indicate the locations of predicted binding motif in the leader sequences for CRISPR loci C and D. The enlarged view of Cas4a promoter shows the locations of the TATA box, BRE, and the third location of the predicted binding motif within the acquisition cassette. B. The palindromic predicted binding motif that is found in three locations within the cluster I-A(2) acquisition cassette: in the identical leader sequences of CRISPR loci C and D as well as in the promoter region of cas4a. Image courtesy of Colin Gauvin.

We also identified this motif directly upstream of CRISPR loci C and D in an identical 231 bp sequence directly adjacent to the first repeat of each locus. These CRISPR loci flank the cluster I-A(2) acquisition cassette, running in opposite directions (Figure 4.1). Though it lacks the hallmarks of a CRISPR locus leader,[35] this repeated sequence containing the motif suggests that Csa3 might act as a transcriptional activator of the CRISPR loci in addition to the acquisition cassette. Further support for this hypothesis comes from the identification of the predicted binding motif directly upstream of CRISPR locus E which is adjacent to the Type III-B(3) cluster.

EMSA Analysis of Potential Binding Sites

After we identified the palindromic motif upstream of the cluster I-A(2) acquisition cassette as well as in the leader sequences of CRISPR loci C, D, and E we attempted to show $cA_4$-stimulated recognition of the sequence via electrophoretic mobility shift assays (EMSA). To examine the binding activity of Csa3 toward the

predicted binding motif and other DNA fragments, we amplified fragments of the *S. solfataricus* genome, including 350 bp sequences centered on the three motif locations (cluster I-A(2) cas4a promoter and CRISPR locus C/D leaders), ~350 bp sequences in the promoter regions of cluster I-A(1) cas4a and cas1, and 175-470 bp sequences from the leader sequences of CRISPR loci B, E, and F.

We found that Csa3 bound each sequence with similar affinity regardless of whether the sequence contains the predicted binding motif or not. Additionally, the inclusion of $cA_4$ did not alter the affinity for any fragment (Figures 4.2-4.5). The measured $K_d$ for the interactions was ~350 +/- 150 nm in each case. Further, the negative control sequence from Lambda phage did lacking the motif, BRE, and TATA binding sites was also recognized with similar affinity, indicating that the observed shift is non-specific. Additionally, a specific binding function lacking a Hill slope variable resulted in a poor fit to the binding data. A good fit upon inclusion of a Hill slope variable suggests that nonspecific binding occurs as multiple Csa3 molecules can associate and dissociate with a single DNA fragment over the course of and EMSA experiment. The resultant Hill coefficient ranged from 2.2-5.5. Interestingly, the incubation temperature did not affect the binding, incubation at 80° C yielded similar results to the shown 37° C incubation experiments. It is possible that the incubation temperature may not show any effect on DNA binding activity because the agarose gel was only run at room temperature.

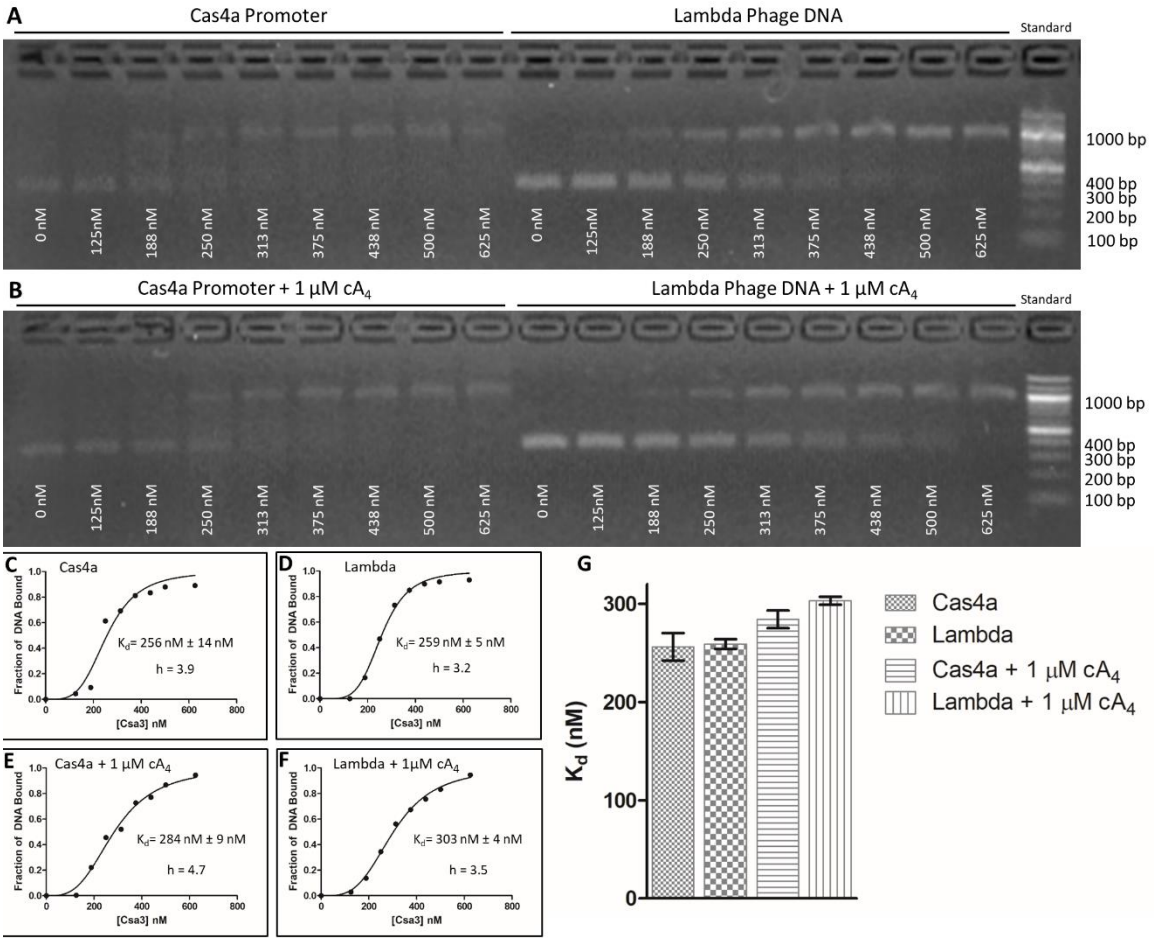


Figure 3.4. EMSA analysis of Csa3 binding to the Cas4a (II) promoter and a Lambda phage DNA control. EMSA experiments were performed with 18 nM DNA and varying concentrations of Csa3 (denoted below the lanes). The last lane of each gel contains a 100bp ladder. EMSA fragments are detailed in Table 3. A. EMSA fragments were incubated without $cA_4$. B. EMSA fragments incubated with 1 μM $cA_4$. C-F. Graphs generated in GraphPad Prism display curves for specific binding with a Hill coefficient (h) fit to binding data generated using ImageJ. G. The bar graph comparing the $K_d$ of each fragment shows that Csa3 binds each with similar affinity and that $cA_4$ has no effect on the binding affinity.

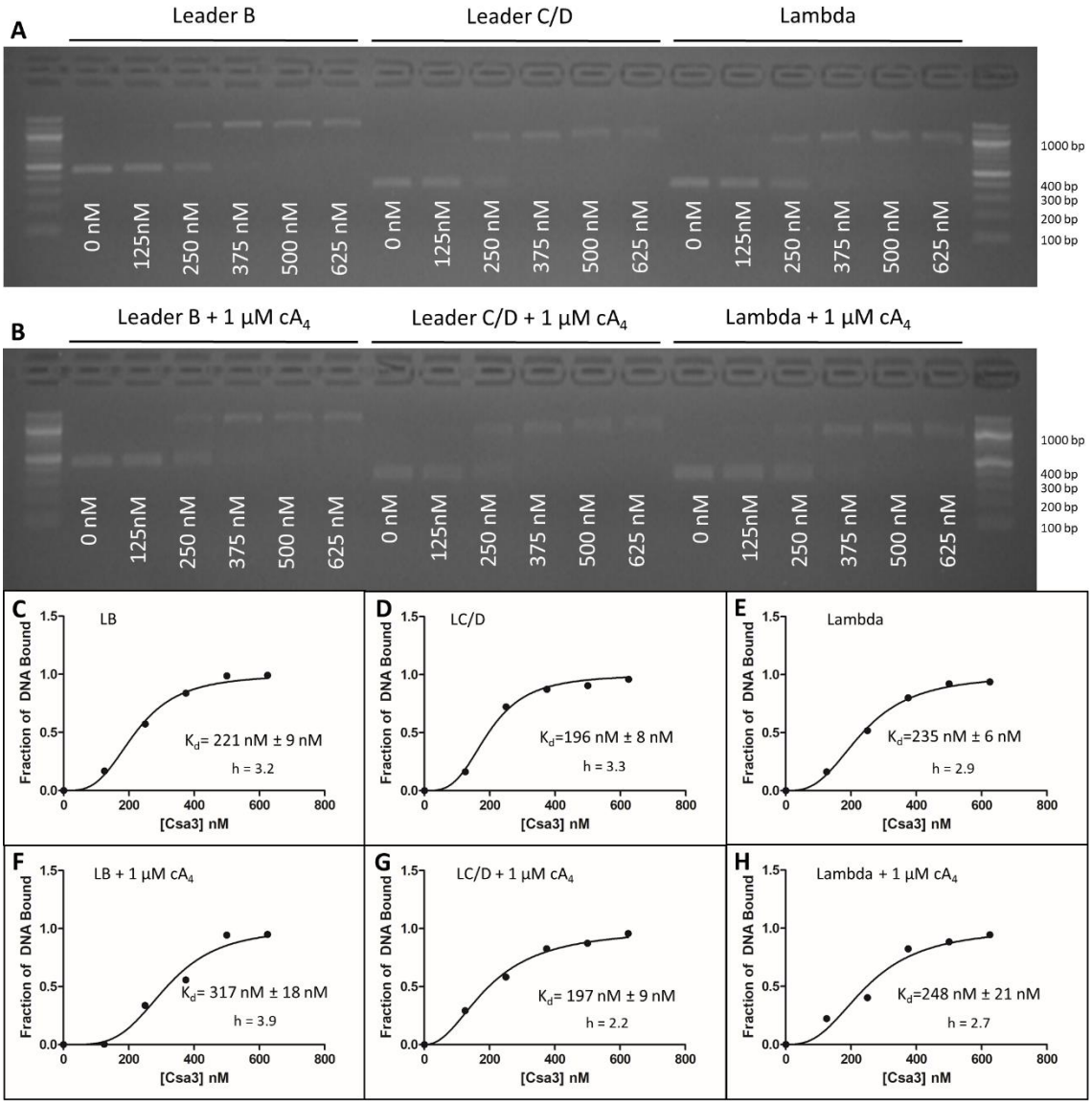


Figure 3.5. EMSA analysis of Csa3 binding to the CRISPR loci B and C/D (identical sequences) Leader sequences. EMSA experiments were performed with 18 nM DNA and varying concentrations of Csa3 (denoted below the lanes). The last lane of each gel contains a 100bp ladder. EMSA fragments are detailed in Table 3. A. EMSA fragments were incubated without $cA_4$. B. EMSA fragments incubated with 1 µM $cA_4$. C-H. Graphs generated in GraphPad Prism display curves for specific binding with a Hill coefficient (h) fit to binding data generated using ImageJ.

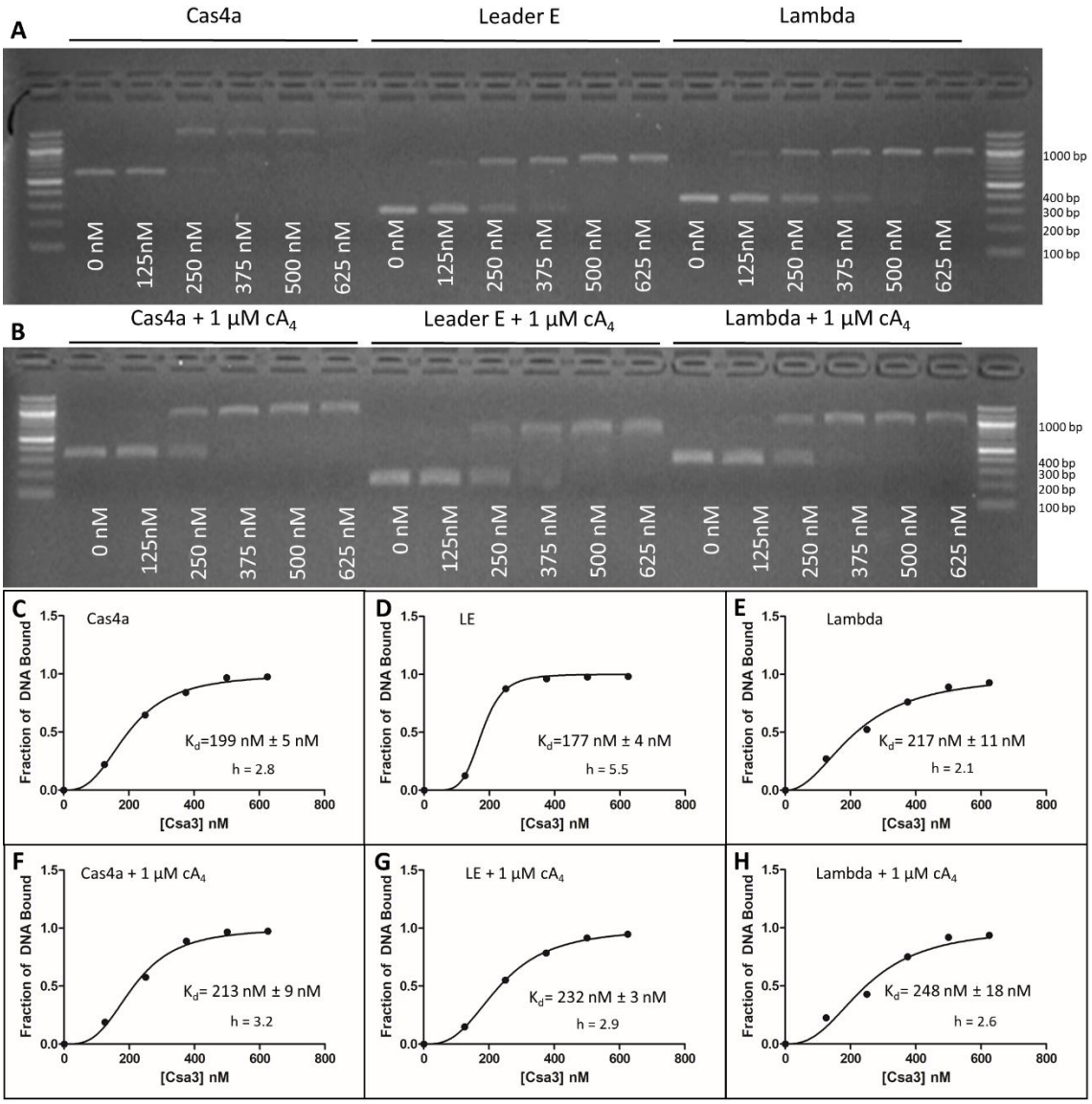


Figure 3.6. EMSA analysis of Csa3 binding to the Cas4a (I) promoter CRISPR locus E Leader sequence l. EMSA experiments were performed with 18 nM DNA and varying concentrations of Csa3 (denoted below the lanes). The last lane of each gel contains a 100bp ladder. EMSA fragments are detailed in Table 3. A. EMSA fragments were incubated without $cA_4$. B. EMSA fragments incubated with 1 μM $cA_4$. C-H. Graphs generated in GraphPad Prism display curves for specific binding with a Hill coefficient (h) fit to binding data generated using ImageJ.

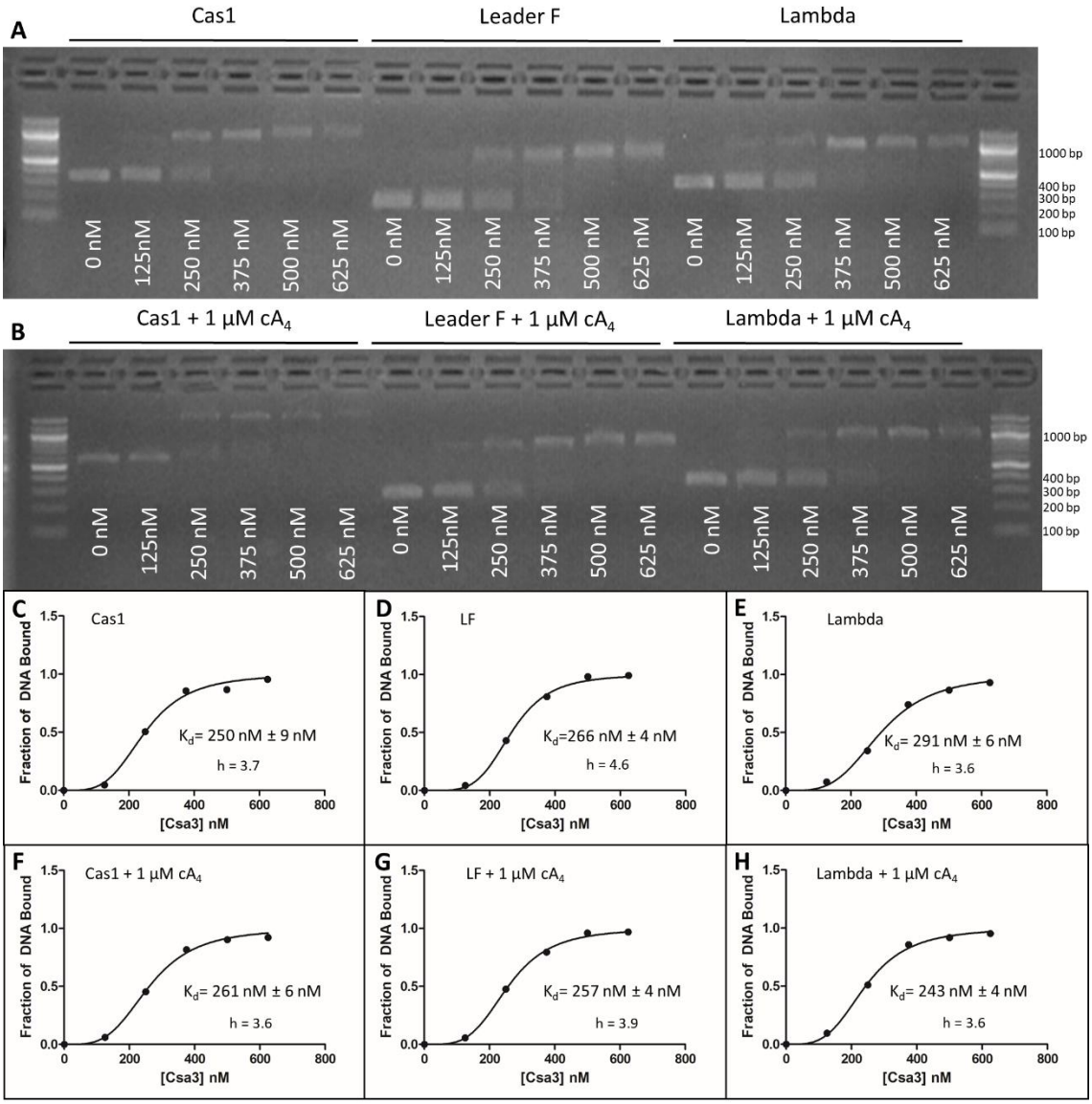


Figure 3.7. EMSA analysis of Csa3 binding to the Cas1 (I) promoter and CRISPR locus F, Leader sequence. EMSA experiments were performed with 18 nM DNA and varying concentrations of Csa3 (denoted below the lanes). The last lane of each gel contains a 100bp ladder. EMSA fragments are detailed in Table 3. A. EMSA fragments were incubated without $cA_4$. B. EMSA fragments incubated with 1 μM $cA_4$. C-H. Graphs generated in GraphPad Prism display curves for specific binding with a Hill coefficient (h) fit to binding data generated using ImageJ.

Csa3 Lacks Ring Nuclease Activity

As with most any signal, the activation of CARF proteins by $cA_4$ needs to be a reversible event or the organism risks negative effects of overactivation, such as the indiscriminate mRNA degradation by Csm6/Csx1.[5,6] Ring nucleases were identified by Athukoralage et al. as a class of CARF domain proteins that specifically degrade $cA_4$ into diadenylate products to degrade the $cA_4$ signal and deactivate indiscriminate ribonuclease activity.[21] Two ring nucleases, Crn1 (Sso2081) and Sso1393 were identified with single turnover rates of $0.23 \pm 0.01$ and $0.024 \pm 0.0004$ $min^{-1}$, respectively.[21,36] These ring nucleases adopt the CARF domain fold, with installation of catalytic residues to degrade bound $cA_4$, yielding 2 molecules of the $A_2$ dinucleotide with terminal 2'-3' cyclic phosphates ($A_2$>P).[21] The presence of this ring nuclease activity in a CARF domain suggested other CARF domain proteins might also harbor ring nuclease activity. Indeed, recent work demonstrates self-limiting CRISPR ribonuclease activity for Csm6 and Csx1, in which the CARF domain slowly degrades the $cA_4$ signal, deactivating the RNase activity of the C-terminal HEPN nuclease domain.[23,37,38] In addition to host ring nucleases, virus-encoded ring nucleases, termed "Type III anti-CRISPR (AcrIII)" proteins, were discovered in a variety of archaeal viruses, plasmids, bacteriophages, and prophages that rapidly degrade $cA_4$ to halt the immune response.[36,39] To determine if Csa3 also harbors ring nuclease activity like Csx1/Csm6, we incubated Csa3 with $cA_4$ and analyzed the resulting products via thin layer chromatography.

Initially, presence of density for the full RNA backbone ring in the structure of the Csa3-$cA_4$ complex, suggested that there may not be nuclease activity in the CARF

domain of Csa3 (Figure 3.1). To confirm to the lack of this prevalent characteristic of CARF proteins, we utilized a ring nuclease assay as described by Athukoralge et al.. As a positive control, *A. fulgidis* Csx3, a CARF domain protein that exhibits ring nuclease activity, showed that ring nuclease cleavage of $cA_4$ results in a product that migrates farther than $cA_4$ on a TLC plate. As expected Csa3 did not exhibit ring nuclease activity, even over the full 7 hr incubation. Given similar affinities to $cA_4$, Csa3 may undergo longer term potentiation than other CARF protein that do exhibit ring nuclease activity because it cannot degrade bound $cA_4$ to shut off its transcriptional regulation. Temporal control of Cas3 activity would therefore rely on the equilibrium established by CARF proteins, ring nucleases, and $cA_4$ production.[36]

# DISCUSSION

As our knowledge of the complexity of the CRISPR-Cas immune system grows with the discovery of new systems, proteins, and functions so does the necessity for understanding the signaling systems that prokaryotes use to coordinate an immune response. Here we report the characterization of the *S. solfataricus* Csa3 structure in complex with the potential allosteric activator $cA_4$ with conformational changes that imply altered activity in the C-terminal DNA-binding domain. Further, the analysis of Csa3 DNA-binding activity through EMSA experiments reveals that the putative transcription factor binds nonspecifically to random DNA at a relatively high affinity, suggesting that interaction with other proteins may be required for specific binding or that Csa3 may not bind B-form DNA. Determination of the $cA_4$-bound structure, along with the identification of a potential binding motif upstream of the I-A(2) acquisition cluster and CRISPR loci C/D support the hypothesis for a signaling pathway in which $cA_4$-activated Csa3 stimulates acquisition of spacers and transcription of CRISPR loci to combat late-stage viral infection.

The structure of Csa3 in complex with $cA_4$ provides structure-function support to the experiments that determined that the cyclic oligoadenylate products of the Type III Csm/Cmr complex are the signals that allosterically control CARF domain proteins.[5,6] Not only is $cA_4$ bound in the predicted CARF binding pocket, its presence causes conformational changes in the CARF domain that lead to significant changes in the spatial relationship between the MarR-like wHTH domains, suggesting that the signal allosterically controls the DNA-binding activity of the protein. The presence of density

for the full $cA_4$ backbone ring implied a lack of ring nuclease activity, which was reinforced by the lack of observed activity in the ring nuclease assay.

Csa3 DNA-binding activity may rely on protein-protein interaction, since EMSA experiments revealed that it does not bind preferentially to sequence with the predicted binding motif, whether or not $cA_4$ is present. It does, however, bind nonspecifically to all tested sequences with a similar, relatively low $K_d$ of ~350 nM. The lack of specificity may imply that Csa3 binds to non-B-form DNA or requires other protein-protein interactions to bind to a specific sequence. The predicted binding motif located upstream of cas4a is flanked by a putative TATA Box and BRE sequence, suggesting that either TATA-Binding Protein, Transcription Factor B,[40,41] or both are required to increase the affinity of Csa3 for that particular DNA sequence. Csa3 may serve to guide these proteins to the sequence or help keep them anchored at the location. The unstructured regions of the C-terminal domain (residues 213-237) may be involved in these interactions as they could extend to interact with other proteins at a much greater distance than other structural features of Csa3 would allow for.

Here we propose a pathway for the $cA_4$-mediated transcriptional activation of acquisition genes and CRISPR loci by Csa3. Csa3 lies adjacent to the Cluster I-A(2) acquisition cassette, suggesting that it controls the transcription of acquisition genes, and thus the acquisition of new spacers. Additionally, we identified an 18 bp palindromic motif that lies directly upstream in the putative promoter region for the acquisition cassette. Searching with the motif, other possible binding sites were identified in front

of the nearby CRISPR loci C and D, revealing that Csa3 might also activate the transcription of newly acquired spacers.

We propose that the transcription factor Csa3 in *S. solfataricus* is activated to stimulate spacer acquisition in a late stage infection after the Type I-A systems have failed due to a mutation in the previously recognized protospacer (Figure 6.1). Type I aCascade first interrogates DNA for a PAM sequence which, though a more efficient method than searching for complimentary base-pairing across the whole strand to begin with, makes them susceptible to small mutation in the invading DNA.[8] If a virus that the cell previously had immunity against due to a complimentary spacer acquires a mutation in either the PAM sequence or protospacer region, it can become invisible to the Type-I aCascade. When this occurs, the invading DNA is transcribed into mRNA to begin translation of viral proteins, at which point the Type III Csm/Cmr complexes interrogate the mRNA in the transcription bubble. Since Type III systems do not utilize a PAM sequence, a PAM mutation does not always affect their ability to identify invading nucleic acids. When the Csm/Cmr complex identifies an infection that has progressed to transcription, the Cas10 subunit synthesizes the $cA_4$ signal, which activates the ancillary CARF proteins. When activated by $cA_4$, Csx1 indiscriminately degrades mRNA, both cellular and viral, leading to cell dormancy and, if not deactivated, eventual cell death.[42] This dormancy buys time for the cell to muster an immune response, a process initiated by Csa3. When activated by $cA_4$, Csa3 stimulates transcription of acquisition genes that integrate new spacers into the CRISPR loci in order to regain immunity to the invading virus. Csa3 also activates transcription of the CRISPR loci, meaning the newly acquired

spacers are immediately transcribed and processed into crRNAs to be loaded into new surveillance complexes. The new aCascade complexes now have restored recognition of the invading virus and can clear the viral DNA. With the viral DNA degraded, transcription of viral mRNA ceases and Csm/Cmr complexes halt production of $cA_4$ while ring nucleases clear the rest of the signal. Loss of the $cA_4$ signal deactivates Csx1, restoring normal transcription and translation, and thus cellular metabolism.

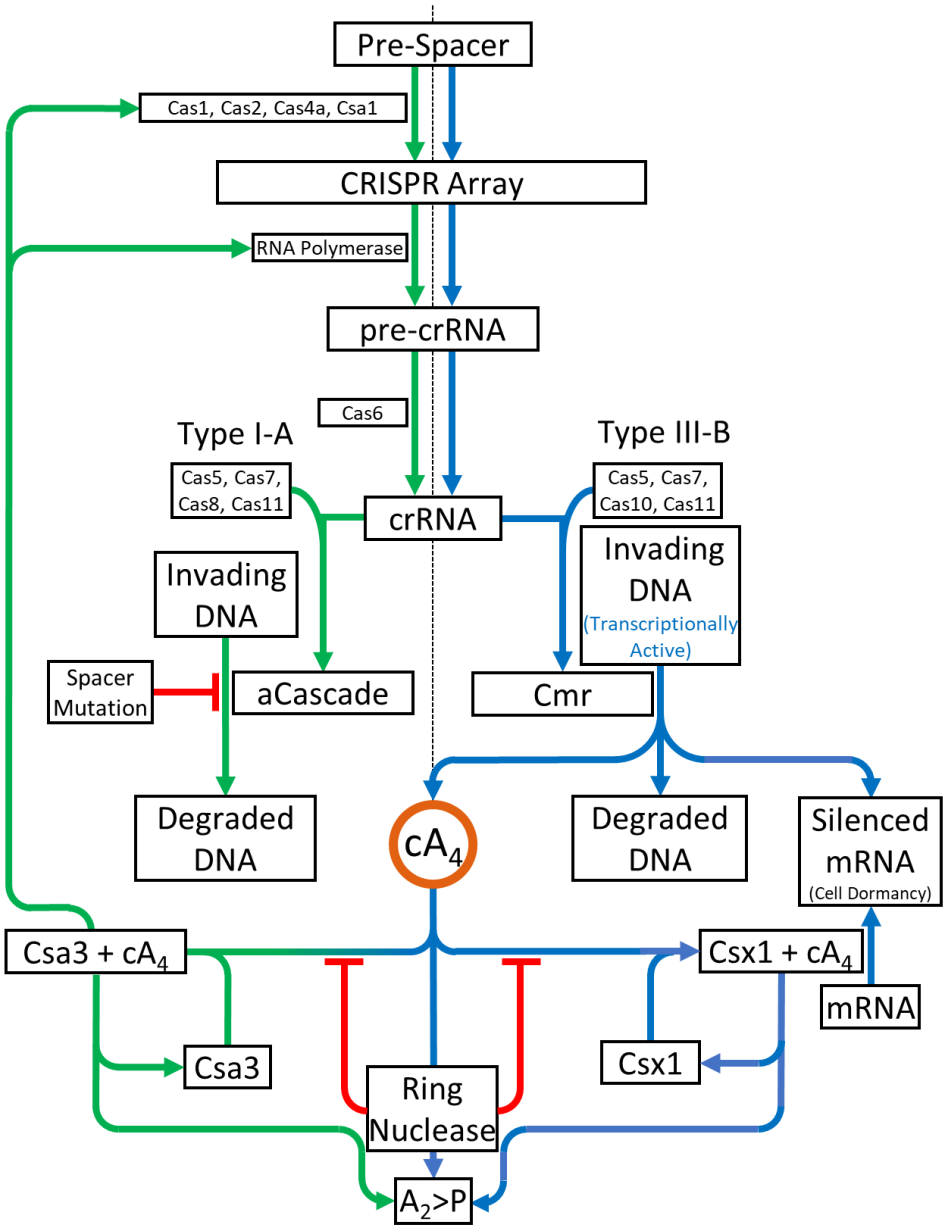


Figure 4.1. Proposed Pathway for $cA_4$-Mediated Transcriptional Activation of Acquisition Genes and CRISPR Loci by Csa3. Blue arrows indicate the pathway taken when the Type I-A systems have lost immunity to a virus due to a spacer mutation. When $cA_4$ is synthesized by the Csm/Cmr complex, it activates Csa3 to stimulate spacer acquisition and crRNA generation, restoring immunity to the Type I-A system. Green arrows indicate the pathway taken to restore immunity to the Type I-A system and clear the infection. Red lines indicate inhibitory processes

# FUTURE WORK

## Csa3 Binding Affinity for $cA_4$ and other Tetranucleotides

Though the crystal structure of the Csa3-$cA_4$ complex shows that $cA_4$ is bound, as expected,[12,16] in the CARF binding pocket, the binding affinity for the interaction is yet unknown, though the dissociation constant is below 1 mM $cA_4$ as we find $cA_4$ in our crystal structure at that concentration. A recent study by Athukoralage et al. revealed that a single strand of recognized viral RNA can cause the production of ~1000 $cA_4$ molecules, which translates to a cellular concentration of ~6 uM $cA_4$, indicating that Csa3 may be active around this concentration of $cA_4$.[36] Knowledge of the binding affinity is essential for comparing the thermodynamics of this interaction to those of other CARF proteins to determine at which $cA_4$ concentration Csa3 is active during an immune response.[36] Data for binding affinities could be obtained through either isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) experiments, or band shift assays with radiolabeled Csa3.

In addition to the binding affinity for $cA_4$, binding data for $A_4$>P and $A_4$ could help determine whether Csa3 specifically binds $cA_4$ or is more promiscuous, like Csm6, which can be activated by $A_4$>P.[5] Further, other cyclic tetranucleotides, such as $cG_4$ may activate Csa3 as well though alternative base-specific interactions, though there is no known source of this molecule in *S. solfataricus*.

## Csa3-Stimulated Spacer Acquisition

*S. islandicus* Csa3 has already been shown to stimulate acquisition of new spacer when overexpressed in the cell,[3] though the same process has not been shown in *S. solfataricus*. In order to show the Csa3-stimulated acquisition of new spacers, the Sso1445 gene can be inserted into a pSeSD vector that contains the *pyrE* gene and then used to transform *S. solfataricus* PH1-16 cells, a P1 variant with an insertion that deactivated the *pyrE* gene.[43] After selecting for the cells using uracil-free defined media, overexpression of Csa3 can be stimulated by addition of arabinose due to the arabinose inducible promoter on the vector.[44] Genomic DNA from Csa3-overexpression cells can be harvested and the first 4-6 spacer sections of each CRISPR loci can be PCR amplified and compared to control cells on via agarose gel electrophoresis. Acquisition of new spacers will appear as a band ~30 bp larger than that of the control cells for each new acquired spacer.

## Csa3 DNA-Binding Activity Under Untested Conditions

Since Csa3 does not show specificity toward the predicted binding motif on its own, other conditions may need to be satisfied for it to bind specifically, such as addition of other proteins or a more biologically relevant temperature. Due to the presence of a putative TATA box and BRE near the predicted binding motif in the I-A(2) cluster, addition of TBP or TFB into the complex may be required to explain the role of Csa3. This could be determined using EMSA experiments similar to those already established, though with the addition of these two proteins. Additionally, if Csa3 serves to increase

the affinity of TFB and/or TBP for the promoter regions containing the motif, experiments such as ITC, SPR, or fluorescence-lifetime spectroscopy could also be used to identify the changes in binding affinity. Further, if other unknown proteins are necessary for specific binding of Csa3 to DNA, it may be possible to add a fragment of biotin-tagged DNA that contains the motif and other binding elements to crude *S. solfataricus* lysate, and then identify proteins that co-purify with it by mass spectrometry.

The DNA-binding activity of Csa3 at the biologically relevant temperature of 80° C may be significantly different than what is observed at room temperature with EMSA experiments. Since docking of Csa3 onto B-form DNA is not possible, it may be that the conserved AT-rich center of the predicted binding motif is melted enough at 80° C to allow Csa3 to bind.

Our determination of the Csa3-$cA_4$ complex provides insight into the structural interactions that occur when $cA_4$ modulates the activity of CARF domain proteins. Additionally, we have developed protocols and techniques for the analysis of Csa3 binding activity that can be used to further explore the signaling pathway that *S. solfataricus* uses to direct a coordinated CRISPR-Cas immune response.

# REFERENCES CITED

(1) Lintner, N. G.; Frankel, K. A.; Tsutakawa, S. E.; Alsbury, D. L.; Copie, V.; Young, M. J.; Tainer, J. A.; Lawrence, C. M. The Structure of the CRISPR-Associated Protein Csa3 Provides Insight into the Regulation of the CRISPR/Cas System. *J Mol Biol* **2011**, *405* (4), 939–955. https://doi.org/10.1016/j.jmb.2010.11.019.

(2) Topuzlu, E.; Lawrence, C. M. Recognition of a Pseudo-Symmetric RNA Tetranucleotide by Csx3, a New Member of the CRISPR Associated Rossmann Fold Superfamily. *RNA Biol*, 2016, *13*, 254–257. https://doi.org/10.1080/15476286.2015.1130209.

(3) Transcriptional Regulator-Mediated Activation of Adaptation Genes Triggers CRISPR de Novo Spacer Acquisition.

(4) Liu, T.; Liu, Z.; Ye, Q.; Pan, S.; Wang, X.; Li, Y.; Peng, W.; Liang, Y.; She, Q.; Peng, N. Coupling Transcriptional Activation of CRISPR-Cas System and DNA Repair Genes by Csa3a in Sulfolobus Islandicus. *Nucleic Acids Res*, 2017, *45*, 8978–8992. https://doi.org/10.1093/nar/gkx612.

(5) Niewoehner, O.; Garcia-Doval, C.; Rostol, J. T.; Berk, C.; Schwede, F.; Bigler, L.; Hall, J.; Marraffini, L. A.; Jinek, M. Type III CRISPR-Cas Systems Produce Cyclic Oligoadenylate Second Messengers. *Nature*, 2017, *548*, 543–548. https://doi.org/10.1038/nature23467.

(6) Kazlauskiene, M.; Kostiuk, G.; Venclovas, C.; Tamulaitis, G.; Siksnys, V. A Cyclic Oligonucleotide Signaling Pathway in Type III CRISPR-Cas Systems. *Science*, 2017, *357*, 605–609. https://doi.org/10.1126/science.aao0100.

(7) Brussaard, C. P. D.; Baudoux, A.-C.; Rodríguez-Valera, F. Marine Viruses. In *The Marine Microbiome: An Untapped Source of Biodiversity and Biotechnological Potential*; Stal, L. J., Cretoiu, M. S., Eds.; Springer International Publishing: Cham, 2016; pp 155–183. https://doi.org/10.1007/978-3-319-33000-6_5.

(8) Hille, F.; Richter, H.; Wong, S. P.; Bratovic, M.; Ressel, S.; Charpentier, E. The Biology of CRISPR-Cas: Backward and Forward. *Cell*, 2018, *172*, 1239–1259. https://doi.org/10.1016/j.cell.2017.11.032.

(9) Sorek, R.; Lawrence, C. M.; Wiedenheft, B. CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu Rev Biochem*, 2013, *82*, 237–266. https://doi.org/10.1146/annurev-biochem-072911-172315.

(10) Lintner, N. G. Crenarchaeal Virus-Host Systems : Structure-Function Studies of Crenarchaeal Viruses and Prokaryotic Adaptive Immunity. *Montana State University*, 2011, *PhD. Thesis*.

(11) Lillestol, R. K.; Shah, S. A.; Brugger, K.; Redder, P.; Phan, H.; Christiansen, J.; Garrett, R. A. CRISPR Families of the Crenarchaeal Genus Sulfolobus: Bidirectional Transcription and Dynamic Properties. *Mol Microbiol*, 2009, *72*, 259–272. https://doi.org/10.1111/j.1365-2958.2009.06641.x.

(12) Lintner, N. G.; Kerou, M.; Brumfield, S. K.; Graham, S.; Liu, H.; Naismith, J. H.; Sdano, M.; Peng, N.; She, Q.; Copie, V.; Young, M. J.; White, M. F.; Lawrence, C. M. Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Associated Complex for

Antiviral Defense (CASCADE). *J Biol Chem*, 2011, *286*, 21643–21656. https://doi.org/10.1074/jbc.M111.238485.

(13) Pyenson, N. C.; Gayvert, K.; Varble, A.; Elemento, O.; Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host & Microbe* **2017**, *22* (3), 343-353.e3. https://doi.org/10.1016/j.chom.2017.07.016.

(14) Haft, D. H.; Selengut, J.; Mongodin, E. F.; Nelson, K. E. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Comput Biol*, 2005, *1*, e60. https://doi.org/10.1371/journal.pcbi.0010060.

(15) Kim, Y. K.; Kim, Y. G.; Oh, B. H. Crystal Structure and Nucleic Acid-Binding Activity of the CRISPR-Associated Protein Csx1 of Pyrococcus Furiosus. *Proteins*, 2013, *81*, 261–270. https://doi.org/10.1002/prot.24183.

(16) Makarova, K. S.; Anantharaman, V.; Grishin, N. V.; Koonin, E. V.; Aravind, L. CARF and WYL Domains: Ligand-Binding Regulators of Prokaryotic Defense Systems. *Front Genet*, 2014, *5*, 102. https://doi.org/10.3389/fgene.2014.00102.

(17) Anantharaman, V.; Makarova, K. S.; Burroughs, A. M.; Koonin, E. V.; Aravind, L. Comprehensive Analysis of the HEPN Superfamily: Identification of Novel Roles in Intra-Genomic Conflicts, Defense, Pathogenesis and RNA Processing. *Biol Direct*, 2013, *8*, 15. https://doi.org/10.1186/1745-6150-8-15.

(18) Sheppard, N. F.; Glover, 3rd, C. V.; Terns, R. M.; Terns, M. P. The CRISPR-Associated Csx1 Protein of Pyrococcus Furiosus Is an Adenosine-Specific Endoribonuclease. *Rna*, 2016, *22*, 216–224. https://doi.org/10.1261/rna.039842.113.

(19) Yan, X.; Guo, W.; Yuan, Y. A. Crystal Structures of CRISPR-Associated Csx3 Reveal a Manganese-Dependent Deadenylation Exoribonuclease. *RNA Biol*, 2015, *12*, 749–760. https://doi.org/10.1080/15476286.2015.1051300.

(20) Zhu, X.; Ye, K. Crystal Structure of Cmr2 Suggests a Nucleotide Cyclase-Related Enzyme in Type III CRISPR-Cas Systems. *FEBS Lett*, 2012, *586*, 939–945. https://doi.org/10.1016/j.febslet.2012.02.036.

(21) Athukoralage, J. S.; Rouillon, C.; Graham, S.; Gruschow, S.; White, M. F. Ring Nucleases Deactivate Type III CRISPR Ribonucleases by Degrading Cyclic Oligoadenylate. *Nature*, 2018, *562*, 277–280. https://doi.org/10.1038/s41586-018-0557-5.

(22) Petit, P.; Xu, X.; Beloglazova, N.; Brown, G.; Savchenko, A.; Yakunin, A. F. Crystal Structure of the CRISPR-Associated Protein SSO1393 from Sulfolobus Solfataricus. *TO BE PUBLISHED*. https://doi.org/10.2210/pdb3qyf/pdb.

(23) Athukoralage, J. S.; Graham, S.; Gruschow, S.; Rouillon, C.; White, M. F. A Type III CRISPR Ancillary Ribonuclease Degrades Its Cyclic Oligoadenylate Activator. *J Mol Biol*, 2019, *431*, 2894–2899. https://doi.org/10.1016/j.jmb.2019.04.041.

(24) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr Purif*, 2005, *41*, 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

(25) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal Biochem*, 1976, *72*, 248–254. https://doi.org/10.1006/abio.1976.9999.

(26) Otwinowski, Z.; Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol*, 1997, *276*, 307–326.
(27) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J Appl Crystallogr*, 2007, *40*, 658–674. https://doi.org/10.1107/s0021889807021206.
(28) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr D Biol Crystallogr*, 2010, *66*, 213–221. https://doi.org/10.1107/S0907444909052925.
(29) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr D Biol Crystallogr*, 2010, *66*, 486–501. https://doi.org/10.1107/S0907444910007493.
(30) Painter, J.; Merritt, E. A. Optimal Description of a Protein Structure in Terms of Multiple Groups Undergoing TLS Motion. *Acta Crystallogr D Biol Crystallogr*, 2006, *62*, 439–450. https://doi.org/10.1107/S0907444906005270.
(31) Chen, V. B.; Arendall, 3rd, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr D Biol Crystallogr*, 2010, *66*, 12–21. https://doi.org/10.1107/S0907444909042073.
(32) DELANO, W. L. The PyMOL Molecular Graphics System. *http://www.pymol.org* **2002**.
(33) Lovell, S. C.; Davis, I. W.; Arendall, 3rd, W. B.; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Calpha Geometry: Phi,Psi and Cbeta Deviation. *Proteins*, 2003, *50*, 437–450. https://doi.org/10.1002/prot.10286.
(34) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M.; Consortium, R. N. A. O. RNA Backbone: Consensus All-Angle Conformers and Modular String Nomenclature (an RNA Ontology Consortium Contribution). *RNA*, 2008, *14*, 465–481. https://doi.org/10.1261/rna.657708.
(35) Alkhnbashi, O. S.; Shah, S. A.; Garrett, R. A.; Saunders, S. J.; Costa, F.; Backofen, R. Characterizing Leader Sequences of CRISPR Loci. *Bioinformatics* **2016**, *32* (17), i576–i585. https://doi.org/10.1093/bioinformatics/btw454.
(36) Athukoralage, J. S.; Graham, S.; Rouillon, C.; Grüschow, S.; Czekster, C. M.; White, M. F. The Dynamic Interplay of Host and Viral Enzymes in Type III CRISPR-Mediated Cyclic Nucleotide Signalling. *bioRxiv* **2020**, 2020.02.12.946046. https://doi.org/10.1101/2020.02.12.946046.
(37) Jia, N.; Jones, R.; Yang, G.; Ouerfelli, O.; Patel, D. J. CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise CA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Mol Cell*, 2019, *75*, 944-956.e6. https://doi.org/10.1016/j.molcel.2019.06.014.

(38) Molina, R.; Stella, S.; Feng, M.; Sofos, N.; Jauniskis, V.; Pozdnyakova, I.; Lopez-Mendez, B.; She, Q.; Montoya, G. Structure of Csx1-COA4 Complex Reveals the Basis of RNA Decay in Type III-B CRISPR-Cas. *Nat Commun*, 2019, *10*, 4302. https://doi.org/10.1038/s41467-019-12244-z.

(39) Athukoralage, J. S.; McMahon, S. A.; Zhang, C.; Grüschow, S.; Graham, S.; Krupovic, M.; Whitaker, R. J.; Gloster, T. M.; White, M. F. An Anti-CRISPR Viral Ring Nuclease Subverts Type III CRISPR Immunity. *Nature* **2020**, *577* (7791), 572–575. https://doi.org/10.1038/s41586-019-1909-5.

(40) Bell, S. D.; Jackson, S. P. Mechanism and Regulation of Transcription in Archaea. *Current Opinion in Microbiology* **2001**, *4* (2), 208–213. https://doi.org/10.1016/S1369-5274(00)00190-9.

(41) Gehring, A. M.; Walker, J. E.; Santangelo, T. J. Transcription Regulation in Archaea. *J. Bacteriol.* **2016**, *198* (14), 1906. https://doi.org/10.1128/JB.00255-16.

(42) Rostol, J. T.; Marraffini, L. A. Non-Specific Degradation of Transcripts Promotes Plasmid Clearance during Type. *Nat Microbiol* **2019**, *4* (4), 656–662. https://doi.org/10.1038/s41564-018-0353-x.

(43) Martusewitsch, E.; Sensen, C. W.; Schleper, C. High Spontaneous Mutation Rate in the Hyperthermophilic Archaeon Sulfolobus Solfataricus Is Mediated by Transposable Elements. *J Bacteriol* **2000**, *182* (9), 2574–2581. https://doi.org/10.1128/jb.182.9.2574-2581.2000.

(44) Peng, N.; Deng, L.; Mei, Y.; Jiang, D.; Hu, Y.; Awayez, M.; Liang, Y.; She, Q. A Synthetic Arabinose-Inducible Promoter Confers High Levels of Recombinant Protein Expression in Hyperthermophilic Archaeon Sulfolobus Islandicus. *Appl. Environ. Microbiol.* **2012**, *78* (16), 5630. https://doi.org/10.1128/AEM.00855-12.

# APPENDIX A

## PYMOL SCRIPTS FOR GENERATION OF FIGURES AND SUPPLEMENTAL MOVIES

OMIT MAP SCRIPT

```
reinitialize

#Windows
#cd \X\Y\Z\pml_Scripts\
load
Model_files\SsCsa3_061319Collection_Current_refine_92_symmetry_40
00000.pdb, SsCsa3

#Mac/Linux
#cd /X/Y/Z/pml_Scripts/
load
Model_files/SsCsa3_061319Collection_Current_refine_92_symmetry_40
00000.pdb, SsCsa3

set map_auto_expand_sym, on
set normalize_ccp4_maps, on
load Csa3_34_B16_101419_11__1_refine_35.mtz, map


set bg_rgb, white
set specular, 0
set light_count, 8
set ambient, 0.15



show_as cartoon, SsCsa3
show_as sticks, chain C
util.cbag chain C


isomesh cA4_map, map.2fofc, 1.0, chain C, carve=1.7

color blue, cA4_map
#color limon, cA4_map
color salmon, chain A
color lightblue, chain B
set ray_shadows, off
set specular, 0
set fog, 0
hide everything, resi 134-212

set_view (\
     0.811870158,   -0.575653017,    0.097417951,\
     0.298006296,    0.265105903,   -0.917011976,\
     0.502054513,    0.773525774,    0.386779815,\
    -0.000000000,    0.000000000, -141.782058716,\
    17.866226196,   44.714614868,   38.565177917,\
   111.782058716,  171.782058716,  -20.000000000 )
```

SUPPLEMENTAL MOVIE SEQUENCE 1 SCRIPT

```
reinitialize

#Windows
#cd \X\Y\Z\pml_Scripts\
load Model_files\Csa3_Morph.pdb, SsCsa3_morph
load Model_files\2wte_wing.pdb, 2wte
load Model_files\cA4_150-_30.pdb, cA4_far
load Model_files\cA4_30-_0.pdb, cA4_near

#Mac/Linux
#cd /X/Y/Z/pml_Scripts/
load Model_files/Csa3_Morph.pdb, SsCsa3_morph
load Model_files/2wte_wing.pdb, 2wte
load Model_files/cA4_150-_30.pdb, cA4_far
load Model_files/cA4_30-_0.pdb, cA4_near


################################################################
#############

set ray_opaque_background, off
set bg_rgb, white

hide everything
set cartoon_flat_sheets = 1
set cartoon_side_chain_helper, off
set cartoon_fancy_helices = 1
set depth_cue = 0
set valence, 0
set cartoon_gap_cutoff=0
set ribbon_as_cylinders, 1
set cartoon_sampling, 14
space rgp

color salmon, SsCsa3_morph and chain A
color salmon, 2wte and chain A
color lightblue, SsCsa3_morph and chain B
color lightblue, 2wte and chain B
color skyblue, chain B and resi 134-212
color tv_red, chain A and resi 134-212
util.cbag cA4_far
util.cbag cA4_near

################################################################
##############
set_view (\
     0.675877333,    0.146095335,    0.722388625,\
     0.140373960,    0.936696172,   -0.320772767,\
```

```
    -0.723522186,    0.318207532,    0.612583816,\
     0.000014424,    0.000000745, -224.191177368,\
   -17.364980698,   -5.121581078,   10.948808670,\
   182.191177368,  266.191253662,  -20.000000000 )

show cartoon, 2wte
show sticks, cA4_far

scene 1, store

hide everything
################################################################
##############
set_view (\
     0.675877333,    0.146095335,    0.722388625,\
     0.140373960,    0.936696172,   -0.320772767,\
    -0.723522186,    0.318207532,    0.612583816,\
     0.000014424,    0.000000745, -224.191177368,\
   -17.364980698,   -5.121581078,   10.948808670,\
   182.191177368,  266.191253662,  -20.000000000 )

show cartoon, SsCsa3_morph
show sticks, cA4_near

scene 2, store

hide everything
################################################################
##############
set_view (\
     0.722388566,   -0.675877631,   -0.146095380,\
    -0.320772737,   -0.140373930,   -0.936696470,\
     0.612584054,    0.723522186,   -0.318207711,\
     0.000000000,    0.000000000,  -97.913238525,\
   -18.073665619,   -1.363563538,   13.752010345,\
    55.913272858,  139.913238525,  -20.000000000 )
clip move, 35

show cartoon, SsCsa3_morph
show sticks, cA4_near

scene 3, store

hide everything
################################################################
##############
set_view (\
     0.722388566,   -0.675877631,   -0.146095380,\
    -0.320772737,   -0.140373930,   -0.936696470,\
     0.612584054,    0.723522186,   -0.318207711,\
     0.000000000,    0.000000000,  -97.913238525,\
   -18.073665619,   -1.363563538,   13.752010345,\
```

```
    55.913272858,  139.913238525,  -20.000000000 )
clip move, 35

show cartoon, SsCsa3_morph
show sticks, cA4_near
show sticks, resi 9+10+11+42+96+97+98+122 and SsCsa3_morph
util.cbay resi 9-11
util.cbap resi 96-98
util.cbao resi 42+122
color violet, resi 9-11 and elem C
color raspberry, resi 42+122 and elem C
hide cartoon, resi 96-98
show cartoon, resi 98 and name CA
show cartoon, resi 96 and name CA
hide cartoon, resi 9-11
show cartoon, resi 11 and name CA
show cartoon, resi 9 and name CA
alter A/40:43/, ss='L'
alter B/40:43/, ss='L'
hide sticks, resi 42 and name C+O+N
hide sticks, resi 122 and sidechain
set cartoon_trace_atoms=1, resi 122 and backbone



scene 4, store

color salmon, SsCsa3_morph and chain A
color salmon, 2wte and chain A
color lightblue, SsCsa3_morph and chain B
color lightblue, 2wte and chain B
hide everything
################################################################
##############

scene 1

mset 1x30 -62 1x1 -62 62x30 -1 1x60 -62 62x30 -1 1x30 1x120 -62
62x60 -1 1x60 -62 62x60 -1 1x60 -62 62x60 -1 1x60 -62 62x60 -1
1x30 -62 62x30
mdo 1: scene 1
mdo 92: scene 2, view=1
mdo 486: scene 2, view=1
mdo 506: scene 3, view=1
mdo 846: scene 4, view=1
mdo 1635: mstop
```

SUPPLEMENTAL MOVIE SEQUENCE 2 SCRIPT

```
reinitialize

#Windows
#cd \X\Y\Z\pml_Scripts\
load Model_files\cA4.pdb, cA4
load Model_files\SsCsa3_chainsAA.pdb, SsCsa3

#Mac/Linux
#cd /X/Y/Z/pml_Scripts/
load Model_files/cA4.pdb, cA4
load Model_files/SsCsa3_chainsAA.pdb, SsCsa3


#####################################################################
#############
set ray_opaque_background, off
set bg_rgb, white

hide everything
set cartoon_flat_sheets=1
set cartoon_side_chain_helper, off
set cartoon_fancy_helices=1
set cartoon_gap_cutoff=0
set ribbon_as_cylinders, 1
set cartoon_sampling, 14

show cartoon, SsCsa3
show sticks, cA4

color salmon, SsCsa3 and chain A
color lightblue, SsCsa3 and chain B


util.cbag cA4
set valence, 0
#################
select 1, /SsCsa3//B/PHE`10/N
select 2, /cA4//C/A`1/O2'
dist Hbond1, 1, 2
delete 1
delete 2
hide labels, HBond1
hide everything, Hbond1
#####
select 1, /SsCsa3//B/PHE`10/N
select 2, /cA4//C/A`2/OP2
dist Hbond2, 1, 2
delete 1
delete 2
```

```
hide labels, HBond2
hide everything, Hbond2
#####
select 1, /SsCsa3//B/ASN`11/ND2
select 2, /cA4//C/A`2/N7
dist Hbond3, 1, 2
delete 1
delete 2
hide labels, HBond3
hide everything, Hbond3
#####
select 1, /SsCsa3//B/GLY`96/N
select 2, /cA4//C/A`2/OP1
dist Hbond4, 1, 2
delete 1
delete 2
hide labels, HBond4
hide everything, Hbond4
#####
select 1, /cA4//C/A`2/OP1
select 2, /SsCsa3//B/GLY`96/N
dist Hbond5, 1, 2
delete 1
delete 2
hide labels, HBond5
hide everything, Hbond5
#####
select 1, /cA4//C/A`3/OP1
select 2, /SsCsa3//B/ARG`98/NH2
dist Hbond6, 1, 2
delete 1
delete 2
hide labels, HBond6
hide everything, Hbond6
#####
select 1, /cA4//C/A`2/N3
select 2, /SsCsa3//B/GLU`122/O
dist Hbond7, 1, 2
delete 1
delete 2
hide labels, HBond7
hide everything, Hbond7
#####
select 1, /SsCsa3//B/ASN`11/N
select 2, /cA4//C/A`2/OP2
dist Hbond8, 1, 2
delete 1
delete 2
hide labels, HBond8
hide everything, Hbond8
#####
select 1, /SsCsa3//B/THR`42/OG1
```

```
select 2, /cA4//C/A`1/N1
dist Hbond9, 1, 2
delete 1
delete 2
hide labels, HBond9
hide everything, Hbond9
#####
color tv_orange, Hbond1
color tv_orange, Hbond2
color tv_orange, Hbond3
color tv_orange, Hbond4
color tv_orange, Hbond5
color tv_orange, Hbond6
color tv_orange, Hbond7
color tv_orange, Hbond8
color tv_orange, Hbond9
```

```
##### Scene 002 #####
set_view (\
     0.722388566,   -0.675877631,   -0.146095380,\
    -0.320772737,   -0.140373930,   -0.936696470,\
     0.612584054,    0.723522186,   -0.318207711,\
     0.000000000,    0.000000000,  -97.913238525,\
   -18.073665619,   -1.363563538,   13.752010345,\
    55.913272858,  139.913238525,  -20.000000000 )
#clip slab, 60m resi 10 and chain B
#clip move, -5
#clip far, 35
#clip near, -10

show sticks, resi 9+10+11+42+96+97+98
show sticks, backbone and resi 122
util.cbay resi 9-11
util.cbap resi 96-98
util.cbao resi 42+122
color violet, resi 9-11 and elem C
color raspberry, resi 42+122 and elem C
hide cartoon, resi 96-98
show cartoon, resi 98 and name CA
show cartoon, resi 96 and name CA
hide cartoon, resi 9-11
show cartoon, resi 11 and name CA
show cartoon, resi 9 and name CA
alter A/40:43/, ss='L'
alter B/40:43/, ss='L'
```

```
hide sticks, resi 42 and name C+O+N
hide sticks, resi 122 and sidechain
set cartoon_trace_atoms=1, resi 122 and backbone

scene 002, store

##### Scene 003 #####
set_view (\
     0.949414611,    0.308843493,   -0.056771595,\
    -0.236585423,    0.584637821,   -0.776030481,\
    -0.206484735,    0.750208914,    0.628134310,\
     0.000016935,    0.000074194,  -54.375297546,\
   -14.342848778,  -20.713277817,    5.987434387,\
  -219.788604736,  328.537689209,  -20.000000000 )
hide everything, chain A
hide everything, resi 4 and chain C
hide everything, name
C5'+C4'+O4'+C3'+O3'+C2'+O2'+C1'+N9+C8+N7+C5+C6+N6+N1+C2+N3+C4 and
resi 3 and chain C

##### Scene 004 #####
set_view (\
     0.953895390,    0.253881484,   -0.160072938,\
    -0.282889158,    0.582372487,   -0.762111366,\
    -0.100267030,    0.772260249,    0.627345443,\
     0.000009138,    0.000061512,  -77.402427673,\
   -12.633566856,  -18.822067261,    8.448163986,\
  -196.762924194,  351.563293457,  -20.000000000 )


hide everything, resi 140-212
util.cbay resi 9-11
util.cbap resi 96-98
util.cbao resi 42+122
color violet, resi 9-11 and elem C
color raspberry, resi 42+122 and elem C
hide cartoon, id 1684-1695
hide cartoon, id 2366-2375
alter A/40:43/, ss='L'
alter B/40:43/, ss='L'
hide sticks, resi 42 and name C+O+N
hide sticks, resi 122 and sidechain
set cartoon_trace_atoms=1, resi 122 and backbone
hide everything, chain A
hide everything, resi 4 and chain C

scene 004, store

##### Scene 006 #####
set_view (\
     0.953895390,    0.253881484,   -0.160072938,\
    -0.282889158,    0.582372487,   -0.762111366,\
```

```
    -0.100267030,    0.772260249,    0.627345443,\
     0.000009138,    0.000061512,  -77.402427673,\
   -12.633566856,  -18.822067261,    8.448163986,\
  -196.762924194,  351.563293457,  -20.000000000 )

show_as cartoon, chain B
hide everything, resi 140-212
color lightblue, SsCsa3 and chain B
show sticks, resi 9+10+11+42+96+97+98 and chain B
show sticks, backbone and resi 122 and chain B
util.cbay resi 9-11
util.cbap resi 96-98
util.cbao resi 42+122
color violet, resi 9-11 and elem C
color raspberry, resi 42+122 and elem C
hide cartoon, id 1684-1695
hide cartoon, id 2366-2375
alter A/40:43/, ss='L'
alter B/40:43/, ss='L'
hide sticks, resi 42 and name C+O+N
hide sticks, resi 122 and sidechain
set cartoon_trace_atoms=1, resi 122 and backbone
hide everything, chain A
hide everything, resi 4 and chain C

show dashes, Hbond1
show dashes, Hbond2
show dashes, Hbond3
show dashes, Hbond4
show dashes, Hbond5
show dashes, Hbond6
show dashes, Hbond7
show dashes, Hbond8
show dashes, Hbond9



scene 006, store

##### 9-11 Scene 007 #####
set_view (\
     0.919013858,    0.050105281,   -0.391030431,\
    -0.352672845,    0.547746241,   -0.758675814,\
     0.176172376,    0.835138321,    0.521056473,\
    -0.000071552,    0.000052889,  -52.686565399,\
   -16.369863510,  -21.193616867,    4.841914177,\
    22.274745941,   83.047668457,  -20.000000000 )
show_as ribbon, chain B
color lightblue, SsCsa3 and chain B
show sticks, resi 9+10+11+42 and chain B
util.cbay resi 9+10+11
util.cbao resi 42
```

```
color violet, resi 9-11 and elem C
color raspberry, resi 42 and elem C
hide ribbon, id 1684-1695
hide cartoon, resi 41-43
show ribbon, resi 40-44 and backbone
hide sticks, resi 42 and name C+O+N
hide everything, chain A
hide everything, resi 4 and chain C

show dashes, Hbond1
show dashes, Hbond2
show dashes, Hbond3
hide dashes, Hbond4
hide dashes, Hbond5
hide dashes, Hbond6
hide dashes, Hbond7
show dashes, Hbond8
show dashes, Hbond9

scene 007, store

##### 96-98 Scene 008 #####
set_view (\
     0.881800234,    0.209772021,   -0.422401190,\
    -0.373424917,    0.857595563,   -0.353660017,\
     0.288062513,    0.469590575,    0.834565043,\
    -0.000026700,    0.000119492,  -56.640579224,\
   -16.158590317,  -18.431001663,    8.095363617,\
    26.223518372,   87.023513794,  -20.000000000 )
show_as ribbon, chain B
color lightblue, SsCsa3 and chain B
show sticks, resi 96+97+98 and chain B
util.cbap resi 96+97+98 and chain B
hide ribbon, id 2366-2375
hide everything, chain A
hide everything, resi 4 and chain C

hide dashes, Hbond1
hide dashes, Hbond2
hide dashes, Hbond3
show dashes, Hbond4
show dashes, Hbond5
show dashes, Hbond6
hide dashes, Hbond7
hide dashes, Hbond8
hide dashes, Hbond9

scene 008, store

##### 122 Scene 009 #####
set_view (\
     0.936732173,    0.153318465,   -0.314680904,\
```

```
    -0.346719265,    0.529976726,   -0.773888290,\
     0.048122808,    0.834031463,    0.549608290,\
     0.000091472,   -0.000105454,  -54.479129791,\
   -21.468069077,  -17.620866776,    1.960267782,\
    24.061717987,   84.861694336,  -20.000000000 )
show_as ribbon, chain B
color lightblue, SsCsa3 and chain B
show sticks, backbone and resi 122 and chain B
util.cbao resi 122 and chain B
color raspberry, resi 122 and elem C
hide everything, chain A
hide everything, resi 4 and chain C

hide dashes, Hbond1
hide dashes, Hbond2
hide dashes, Hbond3
hide dashes, Hbond4
hide dashes, Hbond5
hide dashes, Hbond6
show dashes, Hbond7
hide dashes, Hbond8
hide dashes, Hbond9

scene 009, store
################################################


scene 002
mset 1x1300
mdo 1: scene 002, view=1
mdo 30: scene 002, view=1
mdo 100: scene 004, view=1
mdo 200: rock
mdo 380: rock
mdo 400: scene 004, view=1
mdo 500: scene 006, view=1
mdo 580: scene 006, view=1
mdo 600: scene 007, view=1
mdo 780: scene 007, view=1
mdo 800: scene 008, view=1
mdo 980: scene 008, view=1
mdo 1000: scene 009, view=1
mdo 1180: scene 009, view=1
mdo 1200: scene 006, view=1
mdo 1300: scene 006, view=1
mdo 1300: mstop
```

# DNA-BINDING SCRIPT

```
reinitialize

#Windows
#cd \X\Y\Z\pml_Scripts\
load Model_files\cA4_30-_0.pdb, cA4_near
load Model_files\Csa3_Morph.pdb, SsCsa3
load Model_files\cA4_150-_30.pdb, cA4_far
load Model_files\Seq3_cA4_trans.pdb, cA4_t
load Model_files\2wte_wing.pdb, 2wte
load Model_files\Seq3_Csa3_morphtrans.pdb, Csa3_trans
load Model_files\DNA_t40_short.pdb, DNA

#Mac/Linux
#cd /X/Y/Z/pml_Scripts/
load Model_files/cA4_30-_0.pdb, cA4_near
load Model_files/Csa3_Morph.pdb, SsCsa3
load Model_files/cA4_150-_30.pdb, cA4_far
load Model_files/Seq3_cA4_trans.pdb, cA4_t
load Model_files/2wte_wing.pdb, 2wte
load Model_files/Seq3_Csa3_morphtrans.pdb, Csa3_trans
load Model_files/DNA_t40_short.pdb, DNA



set bg_rgb, white

hide everything
set cartoon_flat_sheets = 1
set cartoon_side_chain_helper, on
set cartoon_sampling, 14
set valence, 0
set cartoon_fancy_helices = 1
set depth_cue = 0
space rgp

#########################################################################
#######

color salmon, SsCsa3 and chain A
color salmon, 2wte and chain A
color salmon, Csa3_trans and chain A
color lightblue, SsCsa3 and chain B
color lightblue, 2wte and chain B
color lightblue, Csa3 and chain B
util.cbag cA4_near
util.cbag cA4_far
util.cbag cA4_t
color lightorange, DNA
color skyblue, chain B and resi 134-212
```

```
color tv_red, chain A and resi 134-212
################################################################
#######
set_view (\
     0.722388625,    0.146095410,   -0.675877631,\
    -0.320772767,    0.936696470,   -0.140373886,\
     0.612584114,    0.318207681,    0.723522186,\
     0.000001101,    0.000001565, -304.493988037,\
   -16.996055603,    5.544666767,   16.098659515,\
   262.493957520,  346.494018555,  -20.000000000 )

hide everything
show cartoon, DNA
show sticks, cA4_far
show cartoon, 2wte

scene 001, store

################################################################
#######
set_view (\
     0.722388625,    0.146095410,   -0.675877631,\
    -0.320772767,    0.936696470,   -0.140373886,\
     0.612584114,    0.318207681,    0.723522186,\
     0.000001101,    0.000001565, -304.493988037,\
   -16.996055603,    5.544666767,   16.098659515,\
   262.493957520,  346.494018555,  -20.000000000 )

hide everything
show cartoon, DNA
show sticks, cA4_near
show cartoon, SsCsa3

scene 002, store

################################################################
#######
set_view (\
     0.988723516,    0.146095410,    0.032888263,\
    -0.127561271,    0.936696470,   -0.326079905,\
    -0.078445047,    0.318207681,    0.944769740,\
     0.000001101,    0.000001565, -304.493988037,\
   -16.996055603,    5.544666767,   16.098659515,\
   262.493957520,  346.494018555,  -20.000000000 )

hide everything
show cartoon, DNA
show sticks, cA4_near
show cartoon, SsCsa3

scene 003, store
```

```
#############################################################################
#######
set_view (\
     0.988723516,    0.146095410,    0.032888263,\
    -0.127561271,    0.936696470,   -0.326079905,\
    -0.078445047,    0.318207681,    0.944769740,\
     0.000001101,    0.000001565, -304.493988037,\
   -16.996055603,    5.544666767,   16.098659515,\
   262.493957520,  346.494018555,  -20.000000000 )

hide everything
show cartoon, DNA
show sticks, cA4_t
show cartoon, Csa3_trans

scene 004, store

#############################################################################
#######
set_view (\
     0.722388625,    0.146095410,   -0.675877631,\
    -0.320772767,    0.936696470,   -0.140373886,\
     0.612584114,    0.318207681,    0.723522186,\
     0.000001101,    0.000001565, -304.493988037,\
   -16.996055603,    5.544666767,   16.098659515,\
   262.493957520,  346.494018555,  -20.000000000 )

hide everything
show cartoon, DNA
show sticks, cA4_t
show cartoon, Csa3_trans

scene 005, store

#############################################################################
#######
scene 001

mset 1x30 -62 # 1x1 -62 62x30 -1 1x30 -62 62x30 -1 1x60 # 1x30 -
62 62x30 -1 1x30 -62 62x30 -1 # 1x1 -62 62x30 -1 1x30 -62 62x30 -
1 1x60 # 1x30 -62 62x30 -1 1x30 -62 62x60
mdo 1: scene 001
mdo 92: scene 002
mdo 450: scene 003
mdo 850: scene 004
mdo 1215: scene 005
mdo 1578: mstop
```

# IMAGES SCRIPT

```
reinitialize

#Windows
#cd \X\Y\Z\pml_Scripts\
load Model_files\SsCsa3_ABC.pdb, SsCsa3

#Mac/Linux
#cd /X/Y/Z/pml_Scripts/
load Model_files/SsCsa3_ABC.pdb, SsCsa3




#################################################################
#############

set ray_opaque_background, off
set bg_rgb, white

hide everything
set cartoon_flat_sheets = 1
set cartoon_side_chain_helper, off
set cartoon_fancy_helices = 1
set depth_cue = 0
set valence, 0
set ribbon_as_cylinders, 1
set cartoon_sampling, 7
space rgp
set specular, 0
set ray_shadows, 0
set light_count, 8
ambient, 0.15

color salmon, SsCsa3 and chain B
color lightblue, SsCsa3 and chain A
color skyblue, chain A and resi 134-212
color tv_red, chain B and resi 134-212
util.cbag chain C

show cartoon, chain A
show cartoon, chain B
show sticks, chain C

#################################################################
############
run Scripts/Views/Standard_View.pml
turn y, 180
scene 1, store
```

```
#####################################################################
###########
run Scripts/Views/Standard_View_Y-45.pml
turn y, 180
scene 2, store

#####################################################################
###########
run Scripts/Views/Standard_View_X-90.pml
turn z, 180


set cartoon_gap_cutoff=0
show sticks, resi 9+10+11+42+96+97+98+122
util.cbay resi 9-11
util.cbap resi 96-98
util.cbao resi 42+122
color violet, resi 9-11 and elem C
color raspberry, resi 42+122 and elem C
hide cartoon, resi 96-98
show cartoon, resi 98 and name CA
show cartoon, resi 96 and name CA
hide cartoon, resi 9-11
show cartoon, resi 11 and name CA
show cartoon, resi 9 and name CA
alter A/40:43/, ss='L'
alter B/40:43/, ss='L'
hide sticks, resi 42 and name C+O+N
hide sticks, resi 122 and sidechain
#set cartoon_trace_atoms=1, resi 122 and backbone
hide everything, resi 133-212

scene 3, store
```

# CUTAWAY VIEW SCRIPT

```
set_view (\
     0.949414611,    0.308843493,   -0.056771595,\
    -0.236585423,    0.584637821,   -0.776030481,\
    -0.206484735,    0.750208914,    0.628134310,\
    -0.000037372,   -0.000088682,  -54.373764038,\
   -19.414276123,  -19.503921509,    7.025605202,\
  -219.788604736,  328.537689209,  -20.000000000 )
```

STANDARD VIEW X-90 SCRIPT

```
set_view (\
     0.722388566,   -0.675877631,   -0.146095380,\
    -0.320772737,   -0.140373930,   -0.936696470,\
     0.612584054,    0.723522186,   -0.318207711,\
     0.000000000,    0.000000000, -198.494873047,\
   -18.073665619,   -1.363563538,   13.752010345,\
   156.494873047,  240.494873047,  -20.000000000 )
```

STANDARD VIEW Y-90 SCRIPT

```
set_view (\
     0.675877333,    0.146095335,    0.722388625,\
     0.140373960,    0.936696172,   -0.320772767,\
    -0.723522186,    0.318207532,    0.612583816,\
     0.000000000,    0.000000000, -198.494873047,\
   -18.073665619,   -1.363563538,   13.752010345,\
   156.494873047,  240.494873047,  -20.000000000 )
```

STANDARD VIEW Y-45 SCRIPT

```
set_view (\
     0.988723338,    0.146095335,    0.032888472,\
    -0.127561212,    0.936696172,   -0.326079965,\
    -0.078445256,    0.318207532,    0.944769561,\
     0.000000000,    0.000000000, -198.494873047,\
   -18.073665619,   -1.363563538,   13.752010345,\
   156.494873047,  240.494873047,  -20.000000000 )
```