

1
2
3 **Mechanism of Filamentation-Induced Allosteric Activation of the SgrAI**
4 **Endonuclease**

5
6 **Smarajit Polley¹, Dmitry Lyumkis^{2*}, Nancy C. Horton^{3,4*}**
7

8
9 ¹Department of Biophysics, Bose Institute, Kolkata, India 700054

10
11 ²Laboratory of Genetics, Salk Institute for Biological Studies, La Jolla, CA, USA 92037

12
13 ³Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA 85721

14
15 ⁴Lead Contact

16
17 ^{*}To whom correspondence should be addressed:

18
19 Nancy C. Horton, phone: (520) 626-3828, email: nhorton@email.arizona.edu

20 Dmitry Lyumkis, phone: (858) 453-4100x1155, email: dlyumkis@salk.edu
21

22
23
24 **SUMMARY**
25

Filament formation by enzymes is increasingly recognized as an important phenomenon with potentially unique regulatory properties and biological roles. SgrAI is an allosterically regulated type II restriction endonuclease that forms filaments with enhanced DNA cleavage activity and altered sequence specificity. Here, we present the cryo-electron microscopy (cryo-EM) structure of the filament of SgrAI in its activated configuration. The structural data illuminate the mechanistic origin of hyper-accelerated DNA cleavage activity and suggests how indirect DNA sequence readout within filamentous SgrAI may enable recognition of substantially more nucleotide sequences than its low-activity form, thereby altering and partially relaxing its DNA sequence specificity. Together, substrate DNA binding, indirect readout, and filamentation simultaneously enhance SgrAI's catalytic activity and modulate substrate preference. This unusual enzyme mechanism may have evolved to perform the specialized functions of bacterial innate immunity in rapid defense against invading phage DNA without causing damage to the host DNA.

INTRODUCTION

Filament formation by non-cytoskeletal enzymes is a newly appreciated phenomenon. Although first shown *in vitro* for particular metabolic enzymes purified from natural sources (Antonini et al., 1966; Gunning, 1965; Kemp, 1971; Kleinschmidt et al., 1969; Meredith and Lane, 1978; Miller et al., 1974; Olsen et al., 1970; Olson and Anfinsen, 1952; Trujillo and Deal, 1977), it was not until many decades later that the biological relevance of this phenomenon was demonstrated (Kim et al., 2010; Shen et al., 2016). At around the same time, filament formation was shown to necessary for the allosteric activation and control of DNA sequence specificity in our system with the endonuclease SgrAI (Park et al., 2010). Finally, large scale screens in cells have identified filament, fiber, and foci formation by many enzymes previously unknown to form such self-assemblies (Narayanaswamy et al., 2009; Noree et al., 2010; Werner et al., 2009). It is now known that more than 20 enzymes from diverse biological pathways form polymeric self-assemblies or filaments (Park and Horton, 2019), yet the role of filamentation in enzyme regulation was not known. To address this gap in knowledge, we have used SgrAI as a model system to investigate fundamental questions regarding enzyme regulation via filament formation, uncovering striking advantages in the form of accelerated rates of enzyme activation, as well as in controlling substrate preference, both of which maximize the *in vivo* function of SgrAI (Barahona et al., 2019; Park et al., 2018a, b). Herein we present a detailed structural investigation to uncover the mechanisms by which filamentation activates the enzymatic activity of SgrAI, and significantly, modulates its DNA sequence specificity.

SgrAI is a sequence-specific type II DNA restriction endonuclease that exhibits both allosteric activation and allosteric modulation of its DNA sequence specificity. Its primary recognition sequence in double stranded DNA (CR|CCGGYG, where R=A or G and Y=C or T, | denotes cleavage site) can both be cleaved and serve as an allosteric activator. When activated, SgrAI cleaves primary sites in DNA over 200-fold faster, but also alters its sequence specificity to include an additional 14 DNA sequences, known as secondary sites (CCCCGGYG or DRCCGGYG, where D=A,G, or T, and underlined nucleotides are those that differ in secondary sites from primary)(Bitinaite and Schildkraut, 2002; Shah et al., 2015). In the absence of DNA, SgrAI is a homodimer composed of two 37 kDa chains, each with a single active site (Park et al., 2010). The enzyme

binds primary or secondary site DNA in a 1:1 ratio, and the ensuing complex is referred to as a DNA-bound dimer (DBD). Since the primary site DNA acts both as an allosteric effector *as well as* a substrate for enzymatic cleavage, it is expected that activated SgrAI should possess at least two DNA binding sites, one for enzymatic cleavage and one for the allosteric effector. Early studies suggested the formation of an activated SgrAI/DNA complex composed of at least two DBDs (Daniels et al., 2003; Wood et al., 2005), thereby providing a functional complex with two DNA binding sites. However, assemblies containing many more DBDs were shown by analytical ultracentrifugation and ion-mobility mass spectrometry (Ma et al., 2013; Park et al., 2010). Negative stain EM revealed filaments of varied lengths with left-handed helical symmetry that we call run-on oligomers (ROO)(Lyumkis et al., 2013).

The ROO filament structure was previously resolved to ~nanometer resolution by cryo-EM and helical reconstruction, and revealed left-handed helical symmetry with approximately 4 DBDs per turn (Lyumkis et al., 2013). This structure inspired a low-resolution mechanistic model for the enzymatic behavior of SgrAI, wherein binding to primary site DNA induces a conformational change that favors ROO filament formation, which in turn stabilizes the activated enzyme state capable of accelerated DNA cleavage (**Figure 1A**). SgrAI DBD bound to a secondary site DNA disfavors the activated conformation, and hence ROO filament formation by SgrAI bound to secondary site does not appreciably occur (**Figure 1B**). This explains why cleavage of secondary site sequences is negligible unless a primary site is present – in sufficient concentration or on the same contiguous DNA – to promote ROO filament assembly (Bitinaite and Schildkraut, 2002; Park et al., 2010; Shah et al., 2015). Filaments formed by SgrAI bound to primary site DNA will drive ROO filament assembly, incorporating DBD-containing secondary sites, activating SgrAI for DNA cleavage on both primary and secondary sites, and thereby altering the sequence specificity of the enzyme (**Figure 1C**).

A full kinetic analysis of SgrAI-mediated cleavage of primary site DNA enabled the estimation of individual rate constants for the major steps in the reaction pathway, such as DBD association into the ROO filament, activated DNA cleavage within the DBD, dissociation of DBD from the ROO filament, and dissociation of cleaved DNA from SgrAI (Park et al., 2018a, b). These studies show that assembly of DBDs into the ROO filament is rate limiting when recognition sites are on separate DNA molecules, but fast when on

the same contiguous DNA. Kinetic simulations show that this rate limiting step underlies the enzyme's ability to sequester potentially damaging secondary site cleavage events so that they occur only on the invading phage DNA, leaving those on the host genome largely untouched (Barahona et al., 2019). The study also shows that DNA cleavage is fast, followed by a relatively slower disassembly of DBDs from the ROO filament. Although slower than DNA cleavage, the rate of disassembly is still sufficiently fast to prevent trapping of cleaved DNA within the filament. Furthermore, simulations of *in vivo* activity comparing filament and non-filament mechanisms show that the filament mechanism exhibits much faster enzyme activation, and also has the ability to better sequester enzyme activity on only substrates of interest, for example on only invading DNA (and not the host DNA). Both of these properties no doubt improve the fitness of the enzyme within its host organism (Barahona et al., 2019).

The previous low-resolution cryo-EM structure showed the general architecture of the ROO filament formed by SgrAI, but left open two important questions: 1) how is the DNA cleavage activity activated within the filament? and 2) how does filament formation regulate the DNA sequence specificity of SgrAI? To address these questions, we sought to determine the structure of activated, filamentous SgrAI to near-atomic resolution. The data reveal how numerous conformational changes in the protein and the bound DNA stabilize the filament, activate DNA cleavage, and control the DNA sequence specificity of SgrAI. In particular, the enzyme may make use of DNA sequence specific stacking interactions (*i.e.* indirect readout of DNA sequence) to control filament formation and hence its activities on the two types of recognition sequences.

RESULTS AND DISCUSSION

Cryo-EM helical analysis of SgrAI ROO filaments resolves the activated enzyme form

To elucidate the molecular determinants of SgrAI activation, we sought to resolve the ROO filament to near-atomic resolution. We prepared ROO filaments from purified, recombinant, his-tagged wild type SgrAI (Shah et al., 2015) bound to a 40-bp oligonucleotide DNA containing a pre-cleaved primary site sequence (PC DNA, see Methods). Cryo-EM micrographs revealed a large variety of differently-sized filaments, as expected for a ROO (**Figure S1A**). Because the filaments were typically limited to $\sim < 10$ DBDs in size, thereby making

it challenging to precisely define their orientations on the cryo-EM micrographs through conventional manual filament tracing, we sought to process the data in a single-particle manner. Template-based particle detection using a small distance between neighboring picks allowed us to select virtually all filamentous regions on the micrographs. 2D and 3D classification – without imposing helical symmetry – allowed us to separate the particle picks into compositionally distinct classes (**Figure 2**, **Figure S1B-C**). Subsequent imposition of helical symmetry facilitated obtaining a high-resolution cryo-EM electrostatic potential map, resolved globally to 3.5 Å, which in turn enabled deriving an atomic model (**Figure 2A**, **Figure S1D-H**). The model is consistent with the final cryo-EM map, and has good geometry and statistics (**Figure S1F**, **Figure S2**, and **Table 1**).

Each DBD constitutes the basic building block of activated and filamentous SgrAI. The enzyme can oligomerize in a run-on manner, from either side along the helical axis (**Figure 2B**). In previous experiments, this property has been shown to specifically benefit the enzyme's ability to quickly sequester phage DNA within the oligomeric form for rapid cleavage (Barahona et al., 2019). Within the filament, SgrAI makes several inter-DBD contacts between asymmetric units. The loop spanning residues 126-134 interfaces with flanking DNA. Electrostatic interactions from charged amino acids within this patch of the protein help maintain the oligomeric enzyme form (**Figure 2C**). This patch also makes up a portion of the allosteric interface of the enzyme. Immediately alongside, loop 55-62 also packs against the flanking DNA, although this interface is primarily maintained by weaker van der Waals interactions with the DNA backbone. Several regions of the N-terminal domain (residues 1-26) make protein-protein interactions between neighboring DBDs within the central filament axis, including clusters of salt bridges that help to facilitate helical packing. Arg84 from DBD_n makes a salt bridge with Asp85 from DBD_{n+1} (**Figure 2D**) (Barahona et al., 2019). Several other charged residues, including Arg11 and Glu8, reside immediately nearby, although their exact contributions cannot be accurately determined at the current resolution. These charged regions from multiple neighboring DBDs, both immediately alongside and at the opposing side of the helix, generally contribute to interface packing along the central helical axis. These data suggest that the interactions made within the central filament axis are specific, and are designed to facilitate oligomerization. The structural organization, combined with previous biochemical and functional data, implies that filament formation is an inherent property of the enzyme (Barahona et al., 2019).

The activated DNA-bound SgrAI enzyme exhibits multiple conformational rearrangements that stabilize the filamentous ROO form

To determine the molecular changes underlying enzyme activation, we compared the structure of the high activity ROO filament to a previously published X-ray structure of a low activity, non-filamentous SgrAI DBD (PDB code 3DVO (Dunten et al., 2008)). Superposition of the two DBDs reveals a $\sim 9^\circ$ rotation of one chain relative to the other (**Figure 3A and Video 1**), with shifts also occurring in the dimeric interface to accommodate this rotation (**Figure 3B**). Such conformational changes allow for favorable interactions to occur within the ROO filament. Specifically, conformational changes place residues 84-87 and 22-34 into position for filament formation (**Figure 3C**). Without such structural changes, the subunits would clash at the interface marked by a red “X”, and thus these residues must facilitate proper helical packing in the filament. Changes in conformation also occur in the bound DNA, and result in favorable interactions between the base pairs flanking the recognition sequence and SgrAI residues (56-57 and 127-134) of a neighboring DBD in the ROO filament. **Figure 3D** shows these interactions, comparing the ROO filament structure, the non-filamentous low-activity DBD, and an idealized B-form DNA. The ROO DNA is bent towards neighboring SgrAI in the ROO filament relative to the idealized B-form DNA, but takes on a different path compared to the DNA in the non-filamentous enzyme assembly. Furthermore, DNA bending allows for interactions with residues 56-57 and 127-134 of the neighboring SgrAI, without steric conflicts. These residues have been shown to be important for enzyme activation, presumably by stabilizing the ROO filament (Shah et al., 2015). Overall, the structural rearrangements observed within the filamentous SgrAI ROO (**Video 1**) prevent unfavorable clashes and promote interactions between neighboring DBDs.

Mechanism of SgrAI hyper-activation for DNA cleavage through stabilization of a second Mg²⁺ binding site

Many nucleases, as well as other phosphoryl transfer enzymes, are dependent on divalent metal ion cofactors such as Mg²⁺ for their activity. SgrAI is no exception, and cleaves nucleic acids in the presence of Mg²⁺, and will also utilize Mn²⁺ or Co²⁺ for catalysis (Dunten et al., 2008; Little et al., 2011). The now classical

two-metal ion mechanism was first proposed for divalent ion-dependent DNA hydrolytic cleavage based on structures of a 3'-5' DNA exonuclease (Beese and Steitz, 1991), and has since been used as a mechanistic model for many metal ion-dependent phosphoryl transfer enzymes (Horton, 2008; Horton and Perona, 2002; Steitz and Steitz, 1993).

Based on the activated cryo-EM reconstruction of filamentous SgrAI, the two-metal ion mechanism, adapted for the enzyme, is shown schematically in **Figure 4A** with experimental structural data supporting the mechanism in **Figure 4B**. In describing both panels, we have utilized the “tense” T and “relaxed” R states – terms adopted based on early enzymology work with hemoglobin (Monod et al., 1965) – to describe the low- and high-activity forms of the enzyme, respectively. In the first panel, SgrAI is in the low activity T state, and only metal ion Site A is occupied. This metal ion ligates both a non-esterified oxygen of the scissile phosphate (SP, **Figure 4A**, panel 1), as well as a water molecule, inducing its deprotonation to hydroxide (blue). The previous X-ray crystal structure of SgrAI bound to uncleaved primary site DNA and Ca^{2+} (PDB file 3DVO (Dunten et al., 2008)), shows this state (**Figure 4B**, panel 1). The next panel of **Figure 4A**, panel 2, shows the shift in conformation to the activated R state, which brings the backbone carbonyl of Thr186 closer to the metal ion binding Site B. The current cryo-EM ROO filament structure of SgrAI bound to Mg^{2+} and primary site DNA (but missing the SP) shows this state (**Figure 4B**, panel 2). In this structure, the Site B Mg^{2+} is unoccupied, presumably due to the absence of the SP. The shift in residues 184-187 is shown by the black arrow. Panel 3 of **Figure 4A** shows the active site immediately prior to the reaction. Once Mg^{2+} binds Site B, it ligates the SP via the same non-esterified oxygen as the Site A Mg^{2+} , as well as the leaving group, the O3'. Although currently there is no direct evidence for ligation between the Site B Mg^{2+} and the O3', it is commonly proposed to occur in two-metal ion mechanisms of other enzymes (Horton and Perona, 2002). The Site B metal ion also ligates a water molecule positioned to donate a proton to the O3' leaving group (cyan, **Figure 4A**, panel 3), thereby stabilizing the otherwise unfavorable negative charge that forms on the O3' as the bond with phosphorus breaks. Panel 3 of **Figure 4B** shows the cryo-EM ROO filament structure in magenta, overlaid onto the Site B Mg^{2+} from a post-catalytic product structure of SgrAI bound to cleaved DNA and residing in the T state (PDB code 3MQY), in blue. The shift in residues 183-188 in the R state brings the backbone carbonyl of Thr186 1.5 Å

closer to the Site B metal ion (blue, **Figure 4B**, panel 3), which would allow for stabilization of the ion either through direct or second shell ligation. Finally, panel 4 of **Figure 4A** shows the predicted product structure after the reaction has occurred. Panel 4 of **Figure 4B** shows the same product structure of 3MQY, with all regions displayed, superimposed on the cryo-EM ROO. Some differences are observed between the product structure of 3MQY (blue, **Figure 4B**, panel 4) and the expected configuration for the product structure in the R-state (**Figure 4A**, panel 4) presumably due to the differences in conformational state, such as in the position of the backbone carbonyl of Thr186, which is too far from the site B Mg^{2+} to for direct coordination. In contrast, the current cryo-EM reconstruction provides the first glimpse into the structural configuration of the enzyme active site in the hyper-activated R state that is competent for rapid DNA cleavage and an improved model for the post-catalytic product state of the hyper-activated form. Together with previous low-activity crystal structures, and building upon prior knowledge of the two-metal ion hydrolytic cleavage mechanism, these snapshots provide the most comprehensive insight into the full mechanism of catalytic cleavage activity for SgrAI.

DNA distortions suggest that ROO-mediated modulation of sequence specificity is mediated by indirect readout of the DNA sequence

To understand how the single base pair difference between primary and secondary site sequences fine-tunes the filamentation and DNA cleavage behavior of SgrAI, we investigated the protein-DNA interactions and DNA structure in the low and high activity conformational states. The low activity state is represented by PDB 3DVO, and the high activity state is represented by the current cryo-EM ROO model. We could find no significant differences in direct readout or in other direct contacts to the 8-bp recognition sequence (**Figure S3**). However, analysis of the DNA structure showed a large rearrangement at the center (4th) base step, between the C4 and G5 nucleotides in CACCCGGTG (**Figure 5A**). The 4 Å rise at this base step is unusually high in both conformations (**Figure 5A right**, the van der Waals rise in B form DNA is 3.4 Å). Such a large rise will likely weaken the stacking energy of these bases. Analysis of RMSD of the 8-bp recognition sequence between the two structures reveals better alignments when superimposing each 4-bp half-site, rather than the full 8-bp

sequence (**Table 2**). Furthermore, calculation of the base overlap areas between neighboring dinucleotide pairs confirmed that the 4th base step undergoes the largest structural change upon transitioning from the low to the high activity state (**Table 3**). Hence, it appears that the DNA accommodates the large conformational change in the SgrAI dimer without disrupting important direct readout contacts between SgrAI and the DNA by allowing each 4-bp half-site to move independently. The weakened stacking and large rise at the center base step allow for this independent movement.

To investigate changes in DNA structure further, the stacking overlap areas were calculated for all base steps of the 8-bp recognition sequence (**Table 3**) in both conformations. Similar DNA structure and stacking areas were found at the first and third base steps (**Table 3, Figure S3A-B**). However, the bases in the second base step, between nucleotides A2 and C3, showed a large shift in positioning (**Table 3, Figure 5B and Video 2**), resulting in a change in stacking area of $\sim 2 \text{ \AA}^2$ (a greater than 2-fold difference). This base step is affected by sequence substitutions in primary site sequences (*i.e.* CRCCGGYG) that are found in two of the fourteen secondary site sequences (*i.e.* CCCCGGYG). Differences in base stacking are likely to result in differences in stacking energy, and hence different degrees of stabilization of the low and high activity conformations.

Mechanistic model of SgrAI activation via filament formation and secondary site cleavage activity

One important issue we seek to fully understand is the mechanism by which primary and secondary site sequences differentially influence SgrAI activation. Secondary sites differ from primary sites by a single base pair, either in the first or second position of the eight-base pair recognition sequence. Some sequence discrimination occurs at the DNA binding step, however SgrAI binds to both types of sites tightly, yet only cleaves secondary sites in the presence of primary (Park et al., 2010; Shah et al., 2015). A working model to rationalize this behavior postulates that SgrAI bound to DNA resides in an equilibrium between a low activity “tense” T state and a high activity “relaxed” R state (**Figure 5C**). Only the high activity R state forms ROO filaments. When the bound DNA has the primary site sequence, the R state is sufficiently populated to result in ROO filament formation (**Figure 5C and Figure 1**). However, when the bound DNA contains the secondary site sequence, the T state is favored to a greater extent, thereby lowering the occupancy of the R state

conformation and hence reducing the propensity to form ROO filaments (**Figure 5C** and **Figure 1**). However, occupancy of the R state is not completely disallowed, since SgrAI bound to secondary site DNA will join filaments formed by SgrAI bound to primary site DNAs (**Figure 1**), thereby stabilizing the R state conformation and activating the enzyme to cleave the secondary site sequences.

In the context of the T and R state model, the X-ray crystal structures of non-filamentous SgrAI bound to DNA (primary or secondary site) represent the T state, and the current cryo-EM ROO filament structure of SgrAI bound to primary site DNA represents the R state. We found that the R state is characterized by a reconfiguration of Thr186 in the enzyme active site, which is predicted to stabilize the site B metal ion (**Figure 4**). Previous data has shown that SgrAI in the R state exhibits rapid DNA cleavage kinetics, and will readily form the ROO filament (**Figure 5C**). The T state has much slower DNA cleavage kinetics and may actually be completely inactive.

The equilibrium between the two states can be estimated by comparing the single turnover DNA cleavage rate constants of SgrAI with primary and secondary site DNAs in the absence of activation and without significant formation of ROO filaments. If only the R state is capable of DNA cleavage (estimated at 0.8 s^{-1}) (Park et al., 2018b), then the observed cleavage rate constant of the unactivated, non-ROO filament form of the enzyme bound to primary site DNA (0.0017 s^{-1}) (Park et al., 2010) can be used to estimate the proportion of SgrAI/DNA complexes in the R and T states. This gives an equilibrium constant that shows that the T state is favored over the R state by 470-fold ($0.8 \text{ s}^{-1}/0.0017 \text{ s}^{-1}$), corresponding to a free energy difference of -3.6 kcal/mol (at 25°C) (blue, **Figure 5D**). Since SgrAI cleaves secondary site DNA much more slowly (Shah et al., 2015), the T state is favored by this complex to a greater extent, giving -4.9 kcal/mol (gold, **Figure 5D**). SgrAI bound to either type of site favors the T state conformation, however when the secondary site sequence is bound, the complex favors the T state conformation by an estimated 1.3 kcal/mol more than when the primary site is bound (**Figure 5D**). This difference in stabilization explains the greater propensity of SgrAI bound to primary sites to form ROO filaments, compared to SgrAI bound to secondary site DNA (**Figure 1**).

Comparison of the R and T state SgrAI structures bound to primary and to secondary DNA should reveal the origin of the differential stabilization. Only a single structure of SgrAI bound to a secondary site has

been determined, that with CCCCGGTG, and in the T state, however no differences in interactions between enzyme and DNA, or in DNA structure, between this and the T state structure with primary site were observed (Little et al., 2011). The current cryo-EM structure shows the R state conformation with primary site DNA. We compared this conformation to those of the T-state, and focused particularly on the interactions and conformation at the 1st and 2nd base pairs, the positions that differ between the primary vs. secondary site sequences. The most significant change was found in the position of the 2nd base pair (**Figure 5B**), resulting in a 50% increase in base stacking overlap area in the R state relative to the T (**Figure 5E**). While the exact energies of base stacking are difficult to calculate, they generally correlate with stacking area. Therefore, this increase in stacking area may be expected to be stabilizing to the R state conformation, relative to the T. Stacking energies measured for different dinucleotide sequences show that the stacking energy of primary site sequences at this base step (AC or GC, blue bars, **Figure 5F**) are more favorable (up to 0.5 kcal/mol) than those of secondary site sequence (CC, gold bar, **Figure 5F**) (Bommarito et al., 2000). Hence, the difference in base stacking and its associated energy may explain at least part of the differential stabilization of the R (vs. T) state by primary and this class of secondary site sequences. A better understanding of the origin in other secondary site sequences (DRCCGGYG, D=G, T, or A) awaits their structure determination in the two conformational states.

Biological role and relationship to other filament forming enzymes

The unusual, allosteric, filament forming mechanism of SgrAI may have evolved due to evolutionary pressure imposed by the relatively large genome of its host, *Streptomyces griseus* (Park et al., 2010). The larger genome results in a greater number of recognition sites, which must be protected from SgrAI-mediated cleavage via methylation by the cognate methyltransferase SgrAI.M. Such pressure to protect these sites from damaging cleavage would favor an increase in the activity of the methyltransferase, and/or a decrease in the activity of the SgrAI endonuclease. We find that the activity of SgrAI is in fact reduced compared to that of similar endonucleases, in that its 8-bp recognition sequence is longer than in typical endonucleases, making it rarer in genomes, and its DNA cleavage rate is remarkably slow in the absence of activation (Park et al., 2010). Activation occurs only during enzyme assembly into ROO filaments when bound to unmethylated primary site

DNA. This is less likely to occur within the host genome, because most primary sites are methylated, but predicted to occur on invading phage DNA containing unmethylated primary sites. In addition to becoming activated upon ROO filament formation, the specificity of SgrAI is also altered and partially relaxed to increase the total number of recognition sequences from 3 to 17. A greater number of potential cleavage sites will result in greater fragmentation of the phage DNA, reducing its capacity for transcription, replication, and repair.

Beyond the well-known cytoskeletal NTPases, filament formation by enzymes is a newly appreciated phenomenon, with the advantages and evolutionary driving forces yet to be fully understood. Proposed purposes of filament formation vary among enzymes and include rapid enzyme activation, sequestration of activity, buffering of activity, and even in functioning as cytoskeletal structures (Barahona et al., 2019; Hunkeler et al., 2018; Liu, 2016; Lynch et al., 2017; O'Connell et al., 2014; O'Connell et al., 2012; Shen et al., 2016; Suresh et al., 2015). In the case of SgrAI, an extensive kinetic study has recently been performed showing that association of SgrAI/DNA complexes into the ROO filament is the rate determining step under most conditions, and is overcome only through high enzyme and DNA concentrations. Owing to local concentration effects, filament formation is expected to occur when two cleavage sites reside on the same contiguous DNA (Barahona et al., 2019; Park et al., 2018a, b). This effect is significant and is responsible for sequestering SgrAI activity on phage DNA and away from the host genome (Barahona et al., 2019). Sequestration of DNA cleavage activity to the activating (*i.e.* phage) DNA is critical, since most secondary sites on the host DNA are not methylated (Morgan, 2019), and can be cleaved by activated SgrAI.

Simulations using the same kinetic parameters determined for SgrAI, but with a theoretical model where only binary oligomeric states are allowed to form, shows that the filament forming mechanism is superior in both speed and sequestration (Barahona et al., 2019). Both the binary and the ROO filament model effectively sequester activated DNA cleavage on phage DNA, with minimal predicted host DNA cleavage, as a result of the slow, rate-limiting enzyme association into higher-order complexes (*i.e.* the binary complex or ROO filaments). However, the ROO filament mechanism is 2-fold faster in DNA cleavage, owing to the multiple ways enzymes can assemble into filaments (*i.e.* at either end), compared to the binary mechanism with only a single manner of association. The binary mechanism can be “sped up” to achieve the same rate of DNA cleavage

as the ROO filament mechanism by increasing the association rate constant for binary assembly formation, however, it must be increased by 4.5-fold to achieve the same fast DNA cleavage kinetics as the ROO filament mechanism (Barahona et al., 2019). This increased association rate constant results in more predicted DNA cleavage of secondary sites on the host DNA, and hence a loss of sequestration of DNA cleavage activity. Speed is likely to be critical to protecting against phage infection by SgrAI (Barahona et al., 2019). Therefore, the ROO filament mechanism may have evolved to meet the opposing requirements of rapid activation and sequestration of activity.

ACKNOWLEDGEMENTS

We thank Yong Zi Tan and Youngmin Jeon for establishing conditions for SgrAI vitrification and Bill Anderson and Jean-Christophe Ducom at The Scripps Research Institute for help with EM data collection and network infrastructure. Molecular graphics and analyses were performed with the USCF Chimera package (supported by NIH P41 GM103311). Research reported in this publication was supported by the National Science Foundation under Grant No. MCB-1410355 (to N.H.) and by the National Institutes of Health Grant No. DP5 OD021396 (to D.L.). SP is supported by the Intermediate Fellowship from DBT Wellcome Trust India Alliance. The EM map and atomic model of activated filamentous SgrAI have been deposited into the EMDB and PDB under accession codes EMD-20015 and 6OBJ respectively.

AUTHOR CONTRIBUTIONS

N.H. prepared and assembled SgrAI and DNA. D.L. collected the cryo-EM data. S.P. processed the cryo-EM data. S.P., N.H., and D.L. built and refined the atomic model; N.H. and D.L. wrote the manuscript; all authors contributed to manuscript editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE TITLES AND LEGENDS

Figure 1. Schematic of differential behavior of SgrAI with primary and secondary site sequences. **A.** SgrAI bound to primary site DNA (cleaved or uncleaved, blue spheres) form ROO filaments with rapid DNA cleavage kinetics. The unfilamentous form cleaves DNA slowly. **B.** SgrAI bound to secondary site DNA only (gold) does not form filaments. **C.** SgrAI bound to secondary site DNA will form ROO filaments with SgrAI bound to primary site DNA, resulting in rapid DNA cleavage of both primary and secondary site sequences.

Figure 2. Structure of activated and oligomeric SgrAI. **(A)** Cryo-EM structure of activated SgrAI ROO filament, reconstructed to 3.5 Å with application of helical symmetry. At right, individual DBDs (DNA bound SgrAI dimers) are colored onto the structure. Scale bar is 100 Å. **(B)** Atomic model of two DBDs, as viewed from the center of the helical axis. The filament can oligomerize in two directions. **(C-D)** Close-up views of interfaces forming inter-DBD contacts are indicated, including **(C)** electrostatic interactions between loop ~126-134 with flanking DNA and **(D)** salt-bridges formed by charged side-chains. See also Figures S1-S2.

Figure 3. Conformational changes facilitate interactions in the ROO filament. **A.** Superposition of the low-activity DBD (3DVO, gray) and the cryo-EM (magenta) structures, with a schematic of the global conformational change. DNA not shown for clarity. **B.** Cartoon rendering of the SgrAI ROO DBD colored by RMSD against the low-activity form. **C.** Protein conformational changes in the ROO prevent steric clashes, which would occur in the low-activity form (marked by “X” in red). **D.** Comparison of flanking DNA between the ROO dimer, the low-activity X-ray structure extended using B-form DNA, and an idealized 40 bp B-form DNA (tan), extended from the 8 bp recognition sequence. In the cryo-EM structure, the DNA of one DBD takes on a distinct path within the filament in order to make contacts with neighboring SgrAI residues 127-134 and 56-57 (green) and prevent steric clashes.

Figure 4. Active site configuration and two-metal ion cleavage mechanism. **A.** Schematic of the two-metal ion mechanism adapted for SgrAI. Panel 1 shows the low activity (T state) conformation where only metal ion Site A is occupied. Panel 2 depicts the R state conformation, based on the cryo-EM ROO filament structure, with the backbone carbonyl of Thr186 shifted closer to stabilize the metal ion of Site B. Panel 3 shows the R state with Site B occupied and just prior to the cleavage reaction. The scissile phosphodiester (SP) containing the bond to be cleaved (red thick line) is indicated. Panel 4 shows the product complex in the R state. **B.** Representative active site structures. Panel 1 shows the

non-filamentous x-ray crystal structure of SgrAI bound to uncleaved primary site DNA and Ca^{2+} occupying Site A (PDB 3DVO), the T state. Panel 2 shows the active site in the cryo-EM ROO filament structure with Mg^{2+} occupying Site A in magenta (the R state), and 3DVO as shown in panel 1B. An arrow shows the shift in residues near Thr186. Panel 3 shows the cryo-EM structure as in Panel 2 (R state, magenta) and the Site B Mg^{2+} (blue), identified in the post-catalytic crystal structure of SgrAI bound to cleaved primary site DNA, in the low activity T state conformation (PDB 3MQY). Panel 4 shows the complete active site arrangement of the post-catalytic product from the low activity T state overlaid on the cryo-EM ROO model. The superposition highlights the large distance traversed by the carbonyl of Thr186 to accommodate the Mg^{2+} -bound R state configuration.

Figure 5. Changes in DNA conformation affect low (T state) and high (R state) conformational energy in DNA sequence dependent manner. A. Change in base stacking in T (represented by the non-ROO structure, PDB 3DVO, colored by atom with carbon atoms gray) and the R state (represented by the cryo-EM ROO filament structure, colored by atom with carbon atoms magenta) at the fourth base step. **Left:** Base stack between base pairs C4:G5' and G5:C4', highlighted in red in the displayed sequence. Blue dotted arrows indicate shift of base upon the T->R change in conformation. **Right:** View of A rotated by 90°. **B.** As in A, but for the second base step. **C.** Equilibrium diagram showing the equilibrium between T and R conformational states of SgrAI/DNA complexes. Only the R state has rapid DNA cleavage kinetics. **D.** Estimate of the equilibrium between the T and R states with primary (T state favored by 3.6 kcal/mol) or secondary site (T state favored by 4.9 kcal/mol) bound to SgrAI. **E.** The base stacking (overlap) area (\AA^2) for the T and R states at the second base step, between bases of nucleotides A2 and C3. **F.** Comparison of stacking energies of possible second base step nucleotides. Primary site sequences (blue) provide more stacking energy than secondary (gold)(Bommarito et al., 2000). See also Figures S3-S4.

TABLES

Table 1.
Cryo-EM data collection, refinement, and validation statistics

EMDB: EMD-20015	
EM data collection and processing	
Microscope	FEI Titan Krios
Voltage (kV)	300
Camera	Gatan K2 Summit
Magnification	22,500
Nominal defocus range (μm)	1.8-3.8
Defocus mean \pm std (μm)	2.6 ± 0.6
Exposure time (s)	12
Number of frames	60
Dose rate (e-/pixel/s)	7.8
Total dose (e-/ \AA^2)	55
Pixel size (\AA)	1.31
Number of micrographs	216
Total extracted particles	31,988
Number of particles in final map	6,864
Symmetry	H
Resolution, FSC 0.143 (\AA)	3.5
Local Resolution Range (\AA)	3.0 – 5.0
Directional Resolution Range (\AA)	3.0 – 4.0
Map sharpening B factor (\AA^2)	-93
Model statistics	
PDB: 6OBJ	
Refinement package (Real space refinement)	Phenix
Model composition	
Protein (residues)	676
DNA (nucleotides)	52
Ligands	2
Map CC	0.82
FSC map-to-model (0.5)	3.5
RMSD [bonds] (\AA)	0.007
RMSD [angles] ($^\circ$)	0.822
All-atom clashscore	5.39
Molprobity score	1.94
Ramachandran plot	
Favored (%)	90.48
Allowed (%)	9.52
Outliers	0
Rotamer outliers (%)	1.42
CaBLAM outliers (%)	4.48
C- β deviations (%)	0
EM-Ringer Score	2.41
Average B-factor, protein (\AA^2)	75.3
Average B-factor, DNA (\AA^2)	99.3

Table 2.**RMSD of DNA between low and high activity conformations of SgrAI/DNA**

Nucleotide of 8 bp Recognition Sequence	RMSD (All Atoms, Duplex) (Å ²)	RMSD (All Atoms, Single Strand) (Å ²)	RMSD (All Atoms, 4 bp Halfsite) (Å ²)
C1	1.44	1.04	0.94
A2	1.05	0.68	0.52
C3	1.07	1.17	1.00
C4	1.14	1.12	0.82
G5	1.03	1.21	0.73
G6	0.62	0.68	0.53
T7	0.82	0.84	0.26
G8	1.16	1.18	0.36
All	1.03	0.99	0.64

Table 3.
Base Overlap Areas

Stacked Bases	Area of Low Activity Conformation, T state (3DVO) (Å ²)	Area of High Activity Conformation, R state (Cryo-EM ROO Filament) (Å ²)	Difference (ROO-3DVO) (Å ²) ^a
1st Base Step CACCGGTG GTGGCCAC			
CA	5.6	5.9	+0.3
TG	0.5	0.05	-0.45
2nd Base Step CACCGGTG GTGGCCAC			
AC	1.6	3.6	+2.0
GT	8.2	8.6	+0.4
3rd Base Step CACCGGTG GTGGCCAC			
CC	6.2	6.6	+0.4
GG	4.5	4.1	-0.3
4th Base Step CACCGGTG GTGGCCAC			
CG	2.8	1.6	-1.2
CG	2.1	1.3	-0.8
Cross strand GG	0	0.9	+0.9

^aNumbers in bold indicate significant differences.

STAR METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, N. C. Horton (nhorton@u.arizona.edu).

This study did not generate new unique reagents.

Plasmid DNA used for protein expression includes a pET21a derivative for SgrAI (derived from a parent plasmid obtained from New England Biolabs, Inc. under a MTA, and available from the Lead Contact's lab under MTA from both the Lead Contact and New England Biolabs, Inc.) and a pBAK derivative constitutively expressing the MspI.M methyltransferase enzyme (obtained from New England Biolabs, Inc. under a MTA, and available only from New England Biolabs, Inc. under MTA with New England Biolabs, Inc.).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The *E. coli* expression system used include BL21 (DE3) strain of *E. coli* carrying plasmids for SgrAI (a pET21a derivative containing 13 addition C-terminal residues containing a TEV protease cleavage site and his-tag: ENLYFQSHHHHHH)(Kong et al., 2000; Shah et al., 2015) and MspI.M (in a derivative of the pBAK plasmid)(Kong et al., 2000; Lin et al., 1989).

METHOD DETAILS

Expression included induction of cultures grown with 50 ug/ml ampicillin and 30 ug/ml kanamycin at 0.6 OD with 0.4 mM IPTG and further growth of the culture at 17°C overnight. After centrifugation, cells are stored at -80°C until used. Cell disruption proceeding using sonication, followed by centrifugation at 10,000 rpm in a Beckman RC2B centrifuge. SgrAI was purified from the cell free lysate using Talon resin chromatography (Clontech, Inc.) and chromatography (GE, Inc.), following manufacturer protocols. Purified SgrAI was

concentrated and stored in single use aliquots at -80°C in buffer containing 50% glycerol. Enzyme purity was assessed using Coomassie blue staining of SDS-PAGE and assessed to at least 99% purity.

DNA preparation

Oligonucleotides were prepared synthetically by a commercial source and purified using C18 reverse phase HPLC. The concentration was measured spectrophotometrically, with an extinction coefficient calculated from standard values for the nucleotides (Fasman, 1975). Equimolar quantities of complementary DNA were annealed by heating to 90°C for 10 minutes at a concentration of 1 mM, followed by slow cooling to room temperature. The sequence of the DNA used in SgrAI/DNA preparations is shown below (red indicates the SgrAI primary recognition sequence):

PC-DNA-top **5'-GATGCGTGGGTCTT**CACA** -3'**
PC-DNA-bottom **3'-CTACGCACCCAGAAGT**GTGGCC**-5'**

Two copies of PC DNA (the duplex formed by annealing of PC-top and PC-bot) self-assemble via annealing of their 5' “overhanging” CCGG sequences to simulate a 40 bp DNA duplex containing a single primary site sequence (shown in red above) after cleavage by SgrAI, however with the exception that it is missing the 5'phosphate at the cleavage site.

EM Sample preparation

SgrAI/DNA samples were prepared using: 30 µl of 6.4 µM SgrAI (in 10 mM Tris-HCl pH 7.8, 300 mM NaCl, 0.1 mM EDTA, 1 mM 2-mercaptoethanol), 2 µl of 420 µM stock PC DNA in H₂O, 1.5 µl of 100 mM Mg(OAc)₂ and incubation at room temperature for 50 min. The final concentrations are 5.8 µM SgrAI, 25 µM PC DNA (4.3:1 ratio of PC DNA:SgrAI), and 4.5 mM MgCl₂. Cryo-EM grids (UltraAuFoil) of this sample were prepared using a manual plunger at 4°C.

QUANTIFICATION AND STATISTICAL ANALYSIS

EM Data collection

Images were recorded on a Titan Krios electron microscope (FEI) equipped with a K2 summit direct detector (Gatan) at 1.31 Å per pixel in counting mode using the Legikon software package (Suloway et al., 2005). Data was acquired using a dose of $\sim 55 \text{ e}^-/\text{Å}^2$ across 60 frames (50 msec per frame) at a dose rate of $\sim 7.8 \text{ e}^-/\text{pix}/\text{sec}$. A total of 216 micrographs were recorded over a single session. All imaging parameters are summarized in *Table 1*.

Image analysis

Preprocessing steps, including frame alignment, CTF estimation, and particle selection, were performed within the Appion pipeline (Lander et al., 2009). Movie frames were aligned using MotionCor2 (Zheng et al., 2017) on 5 by 5 patch squares and using a b-factor of 100. Micrograph CTF estimation was performed using CTFFind4 (Rohou and Grigorieff, 2015). Particles were selected using DoG picker (Voss et al., 2009), with an overlap that did not allow any two picks to be closer than $\sim 150 \text{ Å}$ in distance. This provided sufficient separation within the picks to subsequently enable helical averaging of multiple asymmetric units within extracted particle boxes in the 3D classification and refinement stages. 31,988 particles were extracted at this point. Subsequent steps were performed within the Relion processing package (Scheres, 2012). Reference-free 2D classification was performed to remove any non-filamentous particles. For 3D classification, initially, an asymmetric (C1) classification was employed to separate helical filaments of distinct compositional heterogeneity and to remove bad particles from the data, which did not give rise to an interpretable reconstruction. At the next step, we imposed helical symmetry during 3D classification to classify the particles. After a round of 2D classification, 3D classification without imposition of helical symmetry, and 3D classification with imposition of helical symmetry, 6,894 particles remained. Finally, 3D refinement was performed using a soft-edged mask, and the resulting map was subjected to B-factor sharpening that yielded a final map of the ROO resolved to $\sim 3.5 \text{ Å}$. The final helical symmetry parameters that were used for refinement were 21.6 Å and -86.2° for the rise and twist, respectively. During refinement, we also determined the optimal number of asymmetric units within a box by iteratively varying the number of asymmetric units (and thus how much helical averaging is performed within each box). Based on examination of the FSC curve and visual inspection of the resulting map, this number was

determined to be ~ 7 . Thus, 7-fold helical averaging was performed for each windowed particle. Using this scheme, it is possible that a few of the asymmetric units would appear in the reconstruction more than once, and conversely, that some not at all. However, this was the simplest manner by which to deal with the highly heterogeneous nature of the short, helical fragments that are found in the cryo-EM experiment, while still providing a high-resolution map. The final map was evaluated using Fourier Shell Correlation (FSC) analysis to calculate global and local map resolution [sx_locres.py, implemented within Sparx (Hohn et al., 2007), provided local FSC maps] and the 3DFSC program suite (Tan et al., 2017) to calculate degree of directional resolution anisotropy. All software are summarized in the Key Resources table. The data collection statistics are summarized in Table 1.

Atomic model refinement of 3.5 Å resolution cryo-electron microscopy model

Model building proceeded with real-space refinement after placement of the SgrAI/DNA model derived from X-ray crystallography (PDB 3DVO (Dunten et al., 2008)) using Phenix (Afonine et al., 2018b). Model adjustment and refinement were performed iteratively in Coot (Emsley et al., 2010) and Phenix (Afonine et al., 2018a), and the statistics were examined using Molprobit (Chen et al., 2010) until no further improvements were observed. The final model was also evaluated using FSC analysis against the map and using EMRinger (Barad et al., 2015) to compare the fit of the model backbone into the cryo-EM map. The model statistics showed good geometry and matched the cryo-EM reconstruction (***Figure S1F, Figure S2, and Table 1***).

DNA conformation and base stacking calculations

The software 3DNA (Lu and Olson, 2008) was used to calculate base stacking areas (for all atoms) as well as helical rise.

Structural comparisons-RMSD and alignments

RMSD calculations were performed with the UCSF Chimera package (Pettersen et al., 2004). The “Matchmaker” tool (with Needleman-Wunsch matrix for protein, and “nucleic” for DNA) was used with

structure-based alignment using alpha carbons, phosphorus atoms, backbone atoms only, or all atoms of selected residues of selected chains. Because the scissile phosphate is missing in the EM structure determined here, the three atoms of the scissile phosphate (the phosphorus atom and two non-esterified oxygen atoms) were deleted from the other structures to allow for superposition with all atoms. PyMOL software (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) was used for figures and some alignments, where indicated, using the align command, or the alignment wizard and selected atoms. Analysis of subunit rotation was performed with UCSF Chimera, after first superimposing both chains A of 3DVO (Dunten et al., 2008) and the cryo-EM model using Matchmaker. The match command was used to calculate the rotation angle of chains B relative to each other.

DATA AND CODE AVAILABILITY

The datasets generated during this study are available at the EMDB under code EMD-20015 and the RCSB data base under PDB code 6OBJ.

SUPPLEMENTAL ITEMS TITLES

SUPPLEMENTARY FIGURES

Figure S1. Related to Figure 2. Cryo-EM data, reconstruction and validation. (A) Cryo-EM micrograph of SgrAI. (B) 2D class averages. (C) Reconstructed maps after asymmetric 3D classification showing a class resembling SgrAI (class 1) and two classes with ill-defined features. (D) High-resolution refinement of class 1 from C. In both C and D, a single DBD is highlighted in yellow. (E) Reconstruction from D colored by local resolution, with the full map at left and a cutaway of the center of the filament at right. (F) FSC curves showing half-map resolution (red) and map-to-model resolution (green), both indicating a value of ~ 3.5 Å. 3D-FSC(Tan et al., 2017) isosurfaces are displayed at a threshold of 0.5 within the inset for two perpendicular planar views. (G) segmented DBD map colored by local resolution with (H) the corresponding atomic model.

Figure S2. Related to Figure 2. Correspondence between cryo-EM map and atomic model. Stereo views showing (A) a section of the ROO, (B) an alpha helix (residues 90-122), (C) the active site, with bound Mg²⁺ ions (green), and (D) segmented density of the PC DNA.

Figure S3. Related to Figure 5. Preservation of contacts between SgrAI and the primary recognition sequence in DNA responsible for sequence specific recognition. (A) The low-activity 3DVO structure shown with carbon atoms in gray overlaid with the high-activity ROO filament cryo-EM structure shown with carbon atoms in magenta. Hydrogen bonds indicated by dashed lines. Lys 96 contacts the N3 of G8', Arg 31 makes hydrogen bonds to the O6 and N7 of G8'. (B) Structure at the first base step viewed perpendicular to the base planes. No direct readout contacts are made to the C1 or A2 bases (not shown). Little base stacking (*i.e.* little overlap of the bases in this view) is found between T7' and G8', thought to be part of the indirect readout of the pyrimidine at position 7 of the recognition sequence (Dunten et al., 2008). (C) As in A, but contacts to the C3/G6' base pair. Arg 246 makes hydrogen bonds to the N7 and O6 of G6'. Asp 248 makes a hydrogen bond to the N4 of C3. (D) As in A, but contacts to the C4:G5' bp. Asp 248 makes a hydrogen bond to the N4 of C4, and Arg 249 makes hydrogen bonds to the N7 and O6 of G5'. In all panels, blue, red, and orange refer to nitrogen, oxygen, and phosphorus, respectively.

Figure S4. Related to Figure 5. Base stacking comparison between T and R states. (A) Base Stacking at the first base step. All atoms of the base pair in the background were used in the superposition. Atoms from the cryo-EM ROO filament model representing the R state shown with carbon atoms in magenta. Those of 3DVO, representing the T state, shown with carbon atoms in grey. A shift in the position of the G8' base results in a slight increase in stacking overlap with T7'. (B) As in A, but the third base step. No significant difference is observed.

SUPPLEMENTARY VIDEOS

Supplementary Video 1. Related to Figure 3. Conformational rearrangements between the low- and high-activity SgrAI forms that contribute to helical packing within an ROO

Supplementary Video 2. Related to Figure 5. Conformational rearrangements between the low- and high-activity SgrAI forms that contribute to an increase in base stacking between the 2nd and 3rd base pairs, and thus alteration of its DNA sequence specificity.

REFERENCES

- Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018a). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol* 74, 814-840.
- Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018b). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* 74, 531-544.
- Antonini, E., Brunori, M., Bruzzesi, R., Chiancone, E., and Massey, V. (1966). Association-dissociation phenomena of D-amino acid oxidase. *J Biol Chem* 241, 2358-2366.
- Barad, B.A., Echols, N., Wang, R.Y., Cheng, Y., DiMaio, F., Adams, P.D., and Fraser, J.S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* 12, 943-946.
- Barahona, C.J., Basantes, L.E., Tompkins, K.J., Heitman, D.M., Chukwu, B.I., Sanchez, J., Sanchez, J.L., Ghadirian, N., Park, C.K., and Horton, N.C. (2019). The Need for Speed: Run-On Oligomer Filament Formation Provides Maximum Speed with Maximum Sequestration of Activity. *J Virol* 93.
- Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO Journal* 10, 25-33.
- Bitinaite, J., and Schildkraut, I. (2002). Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease. *Proc Natl Acad Sci U S A* 99, 1164-1169.
- Bommarito, S., Peyret, N., and SantaLucia, J., Jr. (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res* 28, 1929-1934.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12-21.
- Daniels, L.E., Wood, K.M., Scott, D.J., and Halford, S.E. (2003). Subunit assembly for DNA cleavage by restriction endonuclease SgrAI. *J Mol Biol* 327, 579-591.
- Dunten, P.W., Little, E.J., Gregory, M.T., Manohar, V.M., Dalton, M., Hough, D., Bitinaite, J., and Horton, N.C. (2008). The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res* 36, 5405-5416.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.
- Fasman, G.D. (1975). *CRC Handbook of Biochemistry and Molecular Biology*, 3rd edn (Cleveland, OH: CRC).
- Gunning, B.E. (1965). The Fine Structure of Chloroplast Stroma Following Aldehyde Osmium-Tetroxide Fixation. *The Journal of cell biology* 24, 79-93.
- Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. *J Struct Biol* 157, 47-55.
- Horton, N.C. (2008). DNA Nucleases. In *Protein-Nucleic Acid Interactions Structural Biology*, P.A. Rice, Correll, C.C., ed. (Cambridge, UK: The Royal Society of Chemistry), pp. 333-363.
- Horton, N.C., and Perona, J.J. (2002). Making the most of metal ions.[comment]. *Nature Structural Biology* 9, 42-47.
- Hunkeler, M., Hagmann, A., Stutfeld, E., Chami, M., Guri, Y., Stahlberg, H., and Maier, T. (2018). Structural basis for regulation of human acetyl-CoA carboxylase. *Nature* 558, 470-474.
- Kemp, R.G. (1971). Rabbit liver phosphofructokinase. Comparison of some properties with those of muscle phosphofructokinase. *J Biol Chem* 246, 245-252.
- Kim, C.W., Moon, Y.A., Park, S.W., Cheng, D., Kwon, H.J., and Horton, J.D. (2010). Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis. *Proc Natl Acad Sci U S A* 107, 9626-9631.

- Kleinschmidt, A.K., Moss, J., and Lane, D.M. (1969). Acetyl coenzyme A carboxylase: filamentous nature of the animal enzymes. *Science* *166*, 1276-1278.
- Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., *et al.* (2009). Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* *166*, 95-102.
- Little, E.J., Dunten, P.W., Bitinaite, J., and Horton, N.C. (2011). New clues in the allosteric activation of DNA cleavage by SgrAI: structures of SgrAI bound to cleaved primary-site DNA and uncleaved secondary-site DNA. *Acta Crystallogr D Biol Crystallogr* *67*, 67-74.
- Liu, J.L. (2016). The Cytoophidium and Its Kind: Filamentation and Compartmentation of Metabolic Enzymes. *Annu Rev Cell Dev Biol* *32*, 349-372.
- Lu, X.J., and Olson, W.K. (2008). 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* *3*, 1213-1227.
- Lynch, E.M., Hicks, D.R., Shepherd, M., Endrizzi, J.A., Maker, A., Hansen, J.M., Barry, R.M., Gitai, Z., Baldwin, E.P., and Kollman, J.M. (2017). Human CTP synthase filament structure reveals the active enzyme conformation. *Nat Struct Mol Biol* *24*, 507-514.
- Lyumkis, D., Talley, H., Stewart, A., Shah, S., Park, C.K., Tama, F., Potter, C.S., Carragher, B., and Horton, N.C. (2013). Allosteric regulation of DNA cleavage and sequence-specificity through run-on oligomerization. *Structure* *21*, 1848-1858.
- Ma, X., Shah, S., Zhou, M., Park, C.K., Wysocki, V.H., and Horton, N.C. (2013). Structural Analysis of Activated SgrAI-DNA Oligomers Using Ion Mobility Mass Spectrometry. *Biochemistry* *52*, 4373-4381.
- Meredith, M.J., and Lane, M.D. (1978). Acetyl-CoA carboxylase. Evidence for polymeric filament to protomer transition in the intact avian liver cell. *J Biol Chem* *253*, 3381-3383.
- Miller, R.E., Shelton, E., and Stadtman, E.R. (1974). Zinc-induced paracrystalline aggregation of glutamine synthetase. *Archives of biochemistry and biophysics* *163*, 155-171.
- Monod, J., Wyman, J., and Changeux, J.P. (1965). On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol* *12*, 88-118.
- Morgan, R. (2019). Full DNA sequence, and methylation patterns, of the *Streptomyces griseus* genome. Manuscript in preparation.
- Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G.M., O'Connell, J.D., Mirrieles, J., Ellington, A.D., and Marcotte, E.M. (2009). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc Natl Acad Sci U S A* *106*, 10147-10152.
- Noree, C., Sato, B.K., Broyer, R.M., and Wilhelm, J.E. (2010). Identification of novel filament-forming proteins in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. *The Journal of cell biology* *190*, 541-551.
- O'Connell, J.D., Tsechansky, M., Royall, A., Boutz, D.R., Ellington, A.D., and Marcotte, E.M. (2014). A proteomic survey of widespread protein aggregation in yeast. *Mol Biosyst* *10*, 851-861.
- O'Connell, J.D., Zhao, A., Ellington, A.D., and Marcotte, E.M. (2012). Dynamic reorganization of metabolic enzymes into intracellular bodies. *Annu Rev Cell Dev Biol* *28*, 89-111.
- Olsen, B.R., Svenneby, G., Kvamme, E., Tveit, B., and Eskeland, T. (1970). Formation and ultrastructure of enzymically active polymers of pig renal glutaminase. *J Mol Biol* *52*, 239-245.
- Olson, J.A., and Anfinsen, C.B. (1952). The crystallization and characterization of L-glutamic acid dehydrogenase. *J Biol Chem* *197*, 67-79.
- Park, C.K., and Horton, N.C. (2019). Filament Forming Enzymes, a Review. Manuscript in preparation.
- Park, C.K., Sanchez, J.L., Barahona, C., Basantes, L.E., Sanchez, J., Hernandez, C., and Horton, N.C. (2018a). The run-on oligomer filament enzyme mechanism of SgrAI: Part 1. Assembly kinetics of the run-on oligomer filament. *J Biol Chem* *293*, 14585-14598.
- Park, C.K., Sanchez, J.L., Barahona, C., Basantes, L.E., Sanchez, J., Hernandez, C., and Horton, N.C. (2018b). The run-on

- oligomer filament enzyme mechanism of SgrAI: Part 2. Kinetic modeling of the full DNA cleavage pathway. *J Biol Chem* *293*, 14599-14615.
- Park, C.K., Stiteler, A.P., Shah, S., Ghare, M.I., Bitinaite, J., and Horton, N.C. (2010). Activation of DNA cleavage by oligomerization of DNA-bound SgrAI. *Biochemistry* *49*, 8818-8830.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* *25*, 1605-1612.
- Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* *192*, 216-221.
- Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* *180*, 519-530.
- Shah, S., Sanchez, J., Stewart, A., Piperakis, M.M., Cosstick, R., Nichols, C., Park, C.K., Ma, X., Wysocki, V., Bitinaite, J., *et al.* (2015). Probing the Run-On Oligomer of Activated SgrAI Bound to DNA. *PLoS One* *10*, e0124783.
- Shen, Q.-J., Kassim, H., Huang, Y., Li, H., Zhang, J., Li, G., Wang, P.-Y., Ye, F., and Liu, J.-L. (2016). Filamentation of Metabolic Enzymes in *Saccharomyces cerevisiae*. *Journal of Genetics and Genomics* *43*, 393-404.
- Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* *90*, 6498-6502.
- Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., and Carragher, B. (2005). Automated molecular microscopy: the new Legimon system. *J Struct Biol* *151*, 41-60.
- Suresh, H.G., da Silveira Dos Santos, A.X., Kukulski, W., Tyedmers, J., Riezman, H., Bukau, B., and Mogk, A. (2015). Prolonged starvation drives reversible sequestration of lipid biosynthetic enzymes and organelle reorganization in *Saccharomyces cerevisiae*. *Molecular biology of the cell* *26*, 1601-1615.
- Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* *14*, 793-796.
- Trujillo, J.L., and Deal, W.C., Jr. (1977). Pig liver phosphofructokinase: asymmetry properties, proof of rapid association-dissociation equilibria, and effect of temperature and protein concentration on the equilibria. *Biochemistry* *16*, 3098-3104.
- Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., and Carragher, B. (2009). DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* *166*, 205-213.
- Werner, J.N., Chen, E.Y., Guberman, J.M., Zippilli, A.R., Irgon, J.J., and Gitai, Z. (2009). Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc Natl Acad Sci U S A* *106*, 7858-7863.
- Wood, K.M., Daniels, L.E., and Halford, S.E. (2005). Long-range communications between DNA sites by the dimeric restriction endonuclease SgrAI. *J Mol Biol* *350*, 240-253.
- Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* *14*, 331-332.

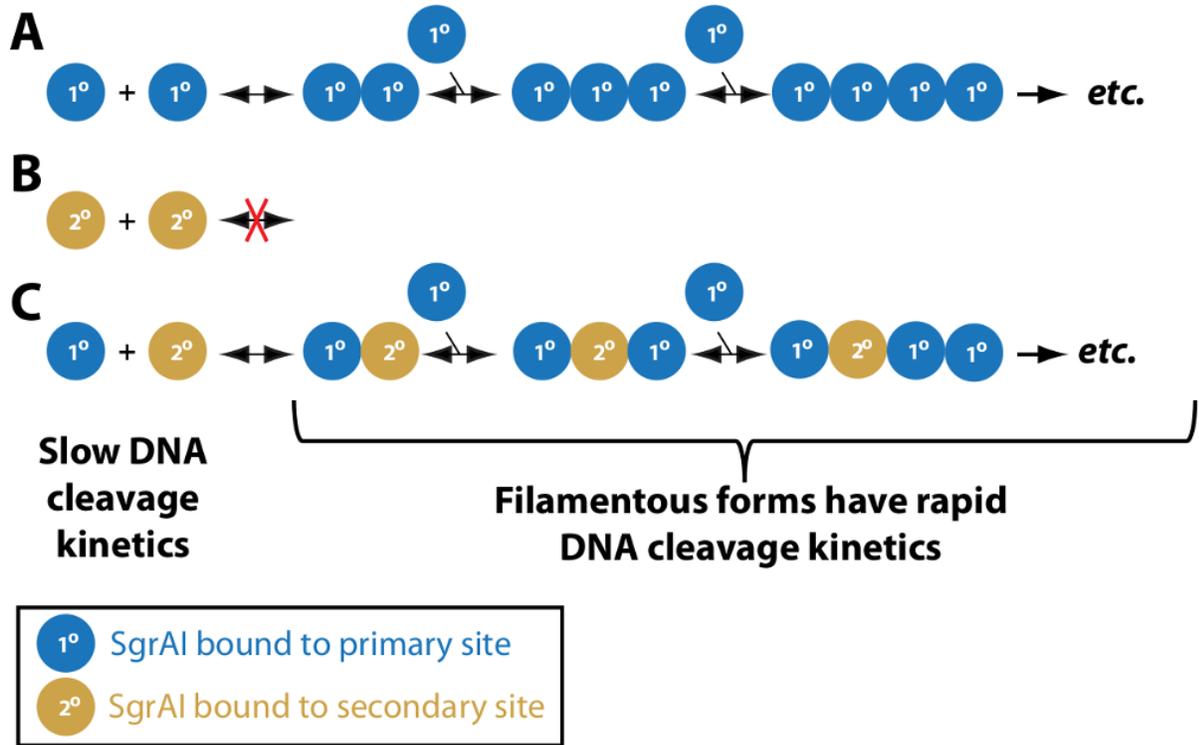


Figure 1. Schematic of differential behavior of SgrAI with primary and secondary site sequences. **A.** SgrAI bound to primary site DNA (cleaved or uncleaved, blue spheres) form ROO filaments with rapid DNA cleavage kinetics. The unfilamentous form cleaves DNA slowly. **B.** SgrAI bound to secondary site DNA only (gold) does not form filaments. **C.** SgrAI bound to secondary site DNA will form ROO filaments with SgrAI bound to primary site DNA, resulting in rapid DNA cleavage of both primary and secondary site sequences.

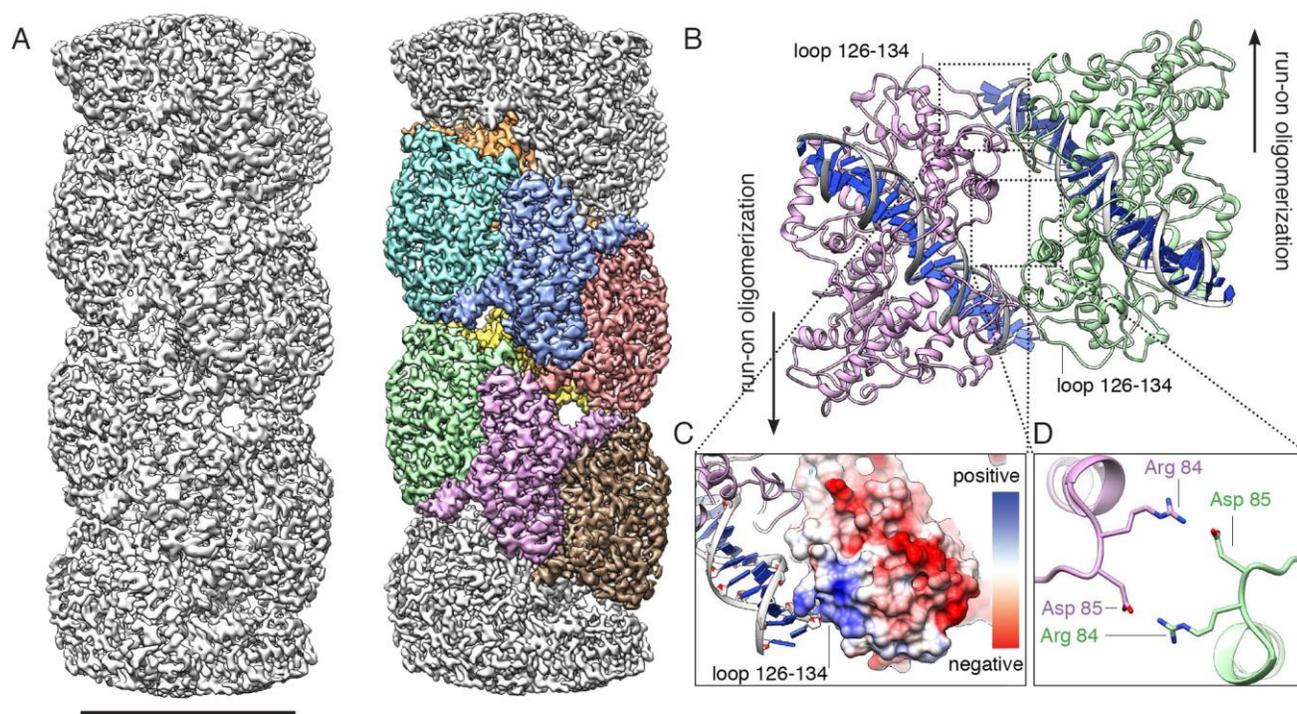


Figure 2. Structure of activated and oligomeric SgrAI. (A) Cryo-EM structure of activated SgrAI ROO filament, reconstructed to 3.5 Å with application of helical symmetry. At right, individual DBDs (DNA bound SgrAI dimers) are colored onto the structure. Scale bar is 100 Å. (B) Atomic model of two DBDs, as viewed from the center of the helical axis. The filament can oligomerize in two directions. (C-D) Close-up views of interfaces forming inter-DBD contacts are indicated, including (C) electrostatic interactions between loop ~126-134 with flanking DNA and (D) salt-bridges formed by charged side-chains. See also Figures S1-S2.

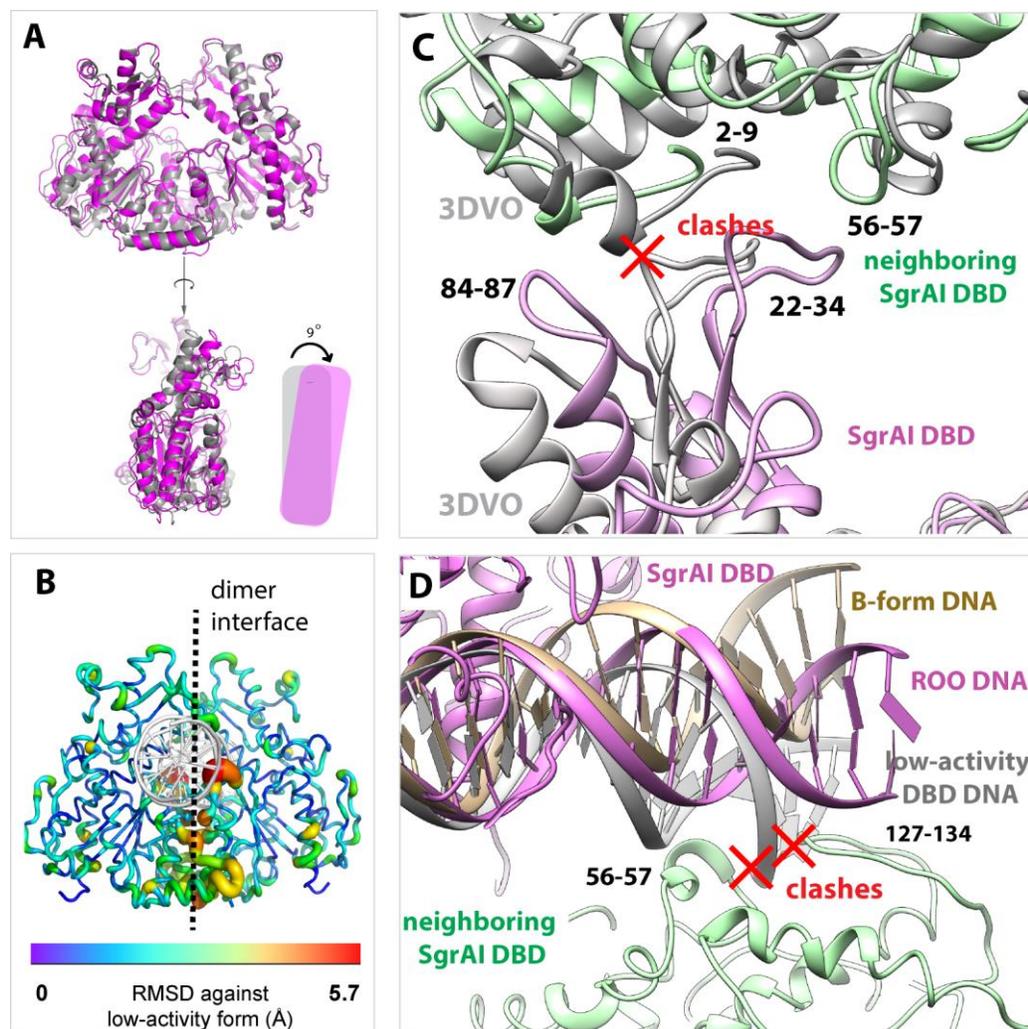


Figure 3. Conformational changes facilitate interactions in the ROO filament. **A.** Superposition of the low-activity DBD (3DVO, gray) and the cryo-EM (magenta) structures, with a schematic of the global conformational change. DNA not shown for clarity. **B.** Cartoon rendering of the SgrAI ROO DBD colored by RMSD against the low-activity form. **C.** Protein conformational changes in the ROO prevent steric clashes, which would occur in the low-activity form (marked by “X” in red). **D.** Comparison of flanking DNA between the ROO dimer, the low-activity X-ray structure extended using B-form DNA, and an idealized 40 bp B-form DNA (tan), extended from the 8 bp recognition sequence. In the cryo-EM structure, the DNA of one DBD takes on a distinct path within the filament in order to make contacts with neighboring SgrAI residues 127-134 and 56-57 (green) and prevent steric clashes.

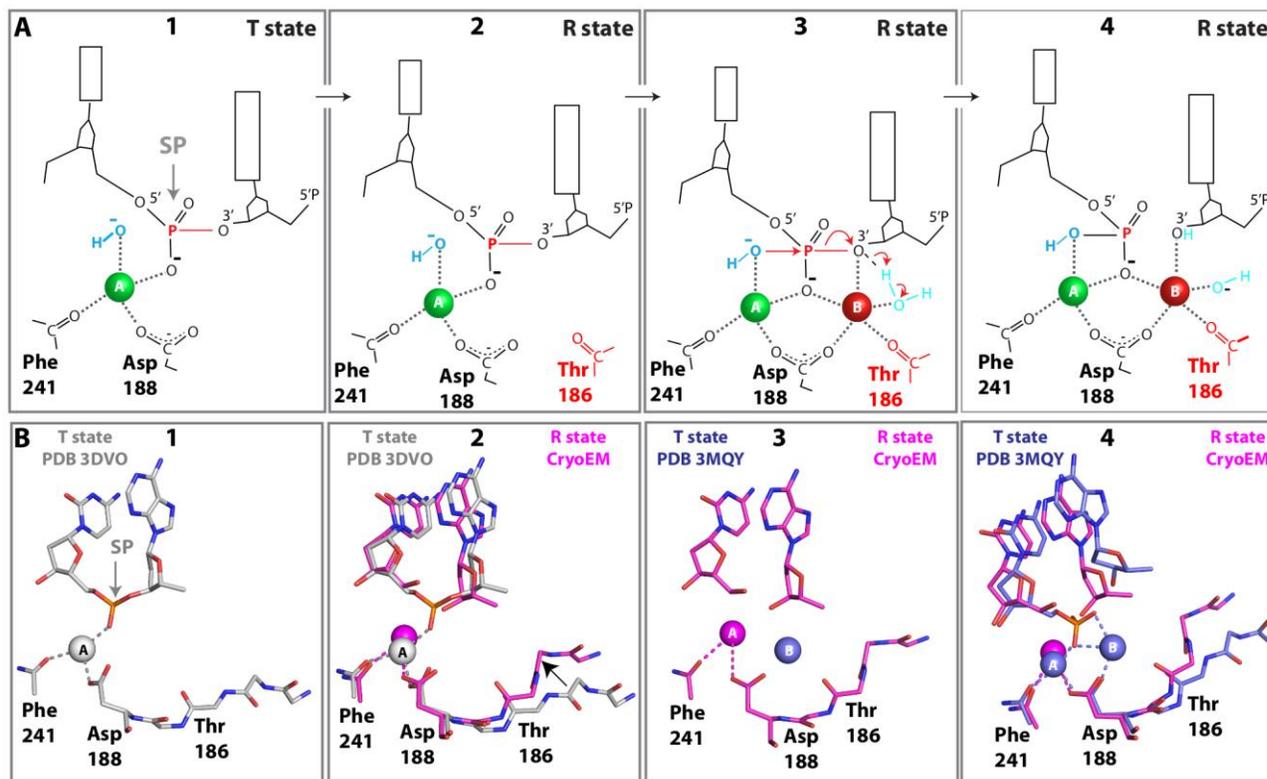


Figure 4. Active site configuration and two-metal ion cleavage mechanism. **A.** Schematic of the two-metal ion mechanism adapted for SgrAI. Panel 1 shows the low activity (T state) conformation where only metal ion Site A is occupied. Panel 2 depicts the R state conformation, based on the cryo-EM ROO filament structure, with the backbone carbonyl of Thr186 shifted closer to stabilize the metal ion of Site B. Panel 3 shows the R state with Site B occupied and just prior to the cleavage reaction. The scissile phosphodiester (SP) containing the bond to be cleaved (red thick line) is indicated. Panel 4 shows the product complex in the R state. **B.** Representative active site structures. Panel 1 shows the non-filamentous x-ray crystal structure of SgrAI bound to uncleaved primary site DNA and Ca^{2+} occupying Site A (PDB 3DVO), the T state. Panel 2 shows the active site in the cryo-EM ROO filament structure with Mg^{2+} occupying Site A in magenta (the R state), and 3DVO as shown in panel 1B. An arrow shows the shift in residues near Thr186. Panel 3 shows the cryo-EM structure as in Panel 2 (R state, magenta) and the Site B Mg^{2+} (blue), identified in the post-catalytic crystal structure of SgrAI bound to cleaved primary site DNA, in the low activity T state conformation (PDB 3MQY). Panel 4 shows the complete active site arrangement of the post-catalytic product from the low activity T state overlaid on the cryo-EM ROO model. The superposition highlights the large distance traversed by the carbonyl of Thr186 to accommodate the Mg^{2+} -bound R state configuration.

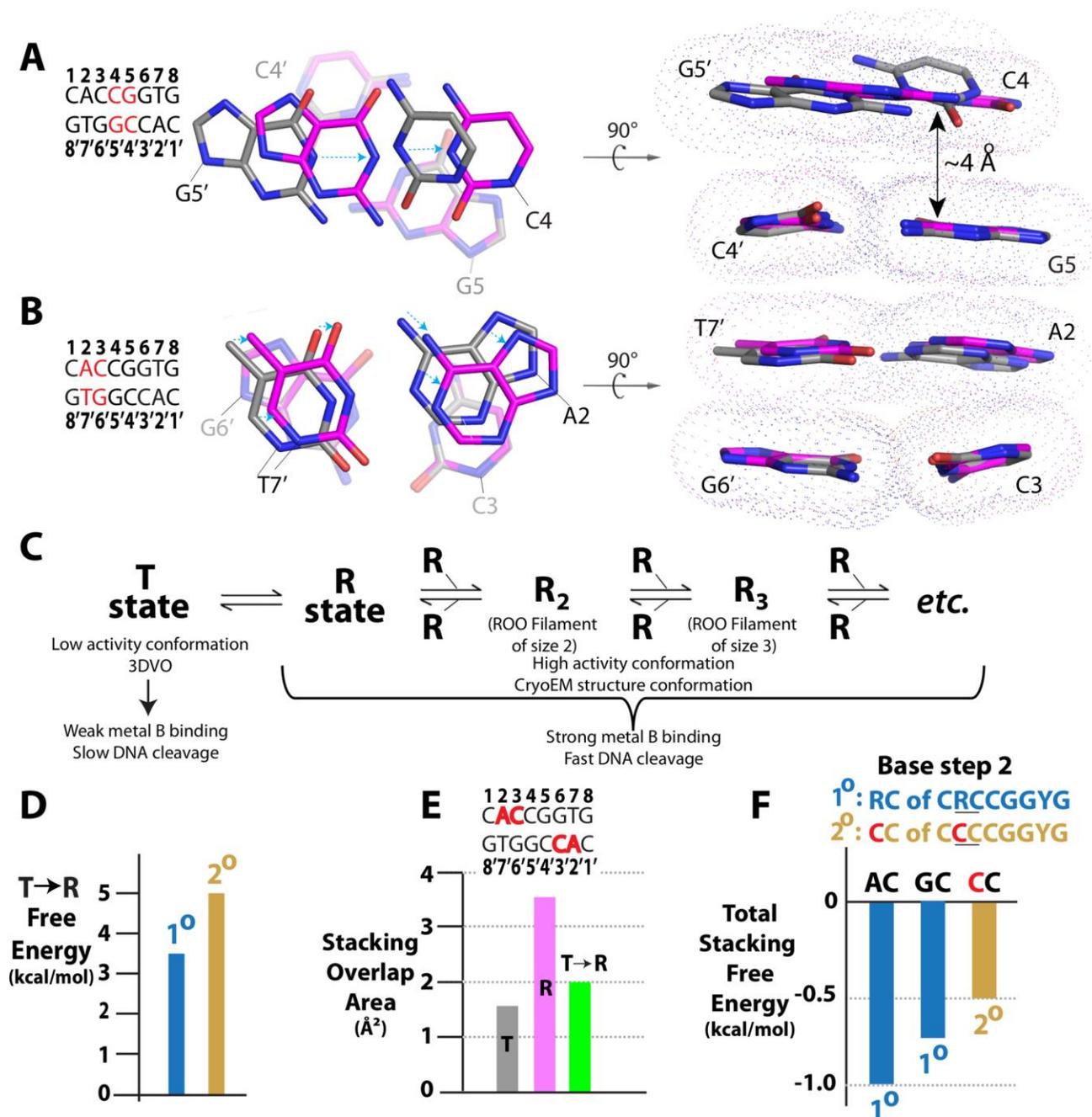
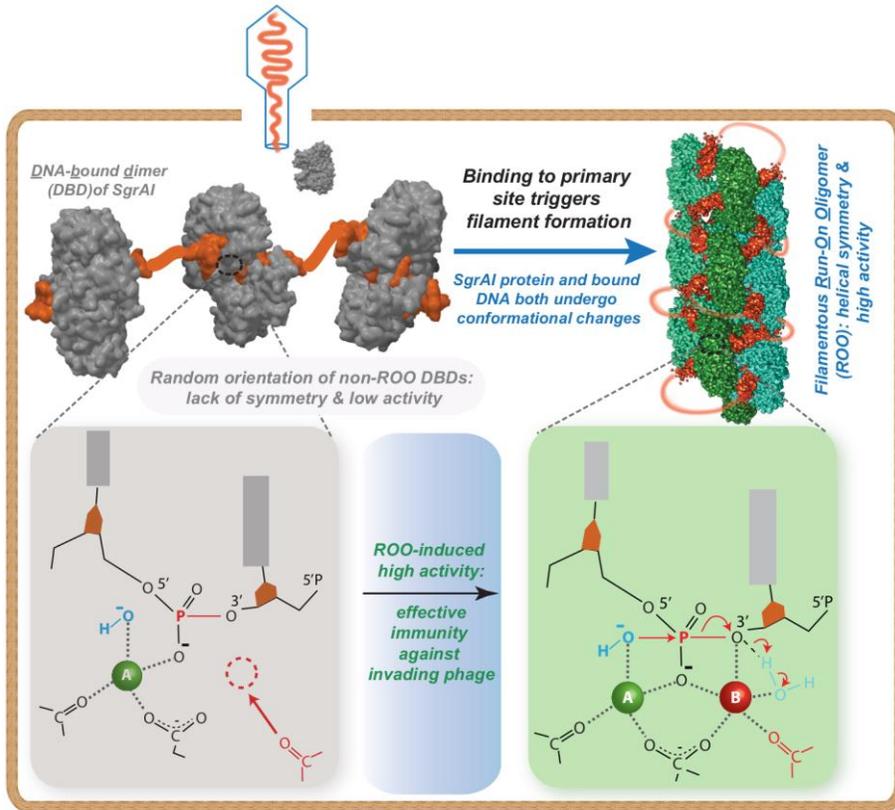


Figure 5. Changes in DNA conformation affect low (T state) and high (R state) conformational energy in DNA sequence dependent manner. **A.** Change in base stacking in T (represented by the non-ROO structure, PDB 3DVO, colored by atom with carbon atoms gray) and the R state (represented by the cryo-EM ROO filament structure, colored by atom with carbon atoms magenta) at the fourth base step. **Left:** Base stack between base pairs C4:G5' and G5:C4', highlighted in red in the displayed sequence. Blue dotted arrows indicate shift of base upon the T→R change in conformation. **Right:** View of A rotated by 90°. **B.** As in A, but for the second base step. **C.** Equilibrium diagram showing

the equilibrium between T and R conformational states of SgrAI/DNA complexes. Only the R state has rapid DNA cleavage kinetics. **D.** Estimate of the equilibrium between the T and R states with primary (T state favored by 3.6 kcal/mol) or secondary site (T state favored by 4.9 kcal/mol) bound to SgrAI. **E.** The base stacking (overlap) area (\AA^2) for the T and R states at the second base step, between bases of nucleotides A2 and C3. **F.** Comparison of stacking energies of possible second base step nucleotides. Primary site sequences (blue) provide more stacking energy than secondary (gold)(Bommarito et al., 2000). See also Figures S3-S4.



Highlights

- A 3.5 Å cryoEM structure of the filamentous form of DNA bound SgrAI.
- Conformational changes play a role in catalytic Mg^{2+} stabilization
- New configuration explains enzyme activation in filamentous state
- Indirect readout of DNA structure associated with SgrAI filamentation and activity

eTOC:

The 3.5 Å Cryo-EM structure of the filament formed by the type II restriction endonuclease SgrAI bound to DNA described in Polley et al. reveals the relevance of indirect readout for enzyme activity and a mechanism for filament-induced activation of DNA cleavage.

Supplementary Information

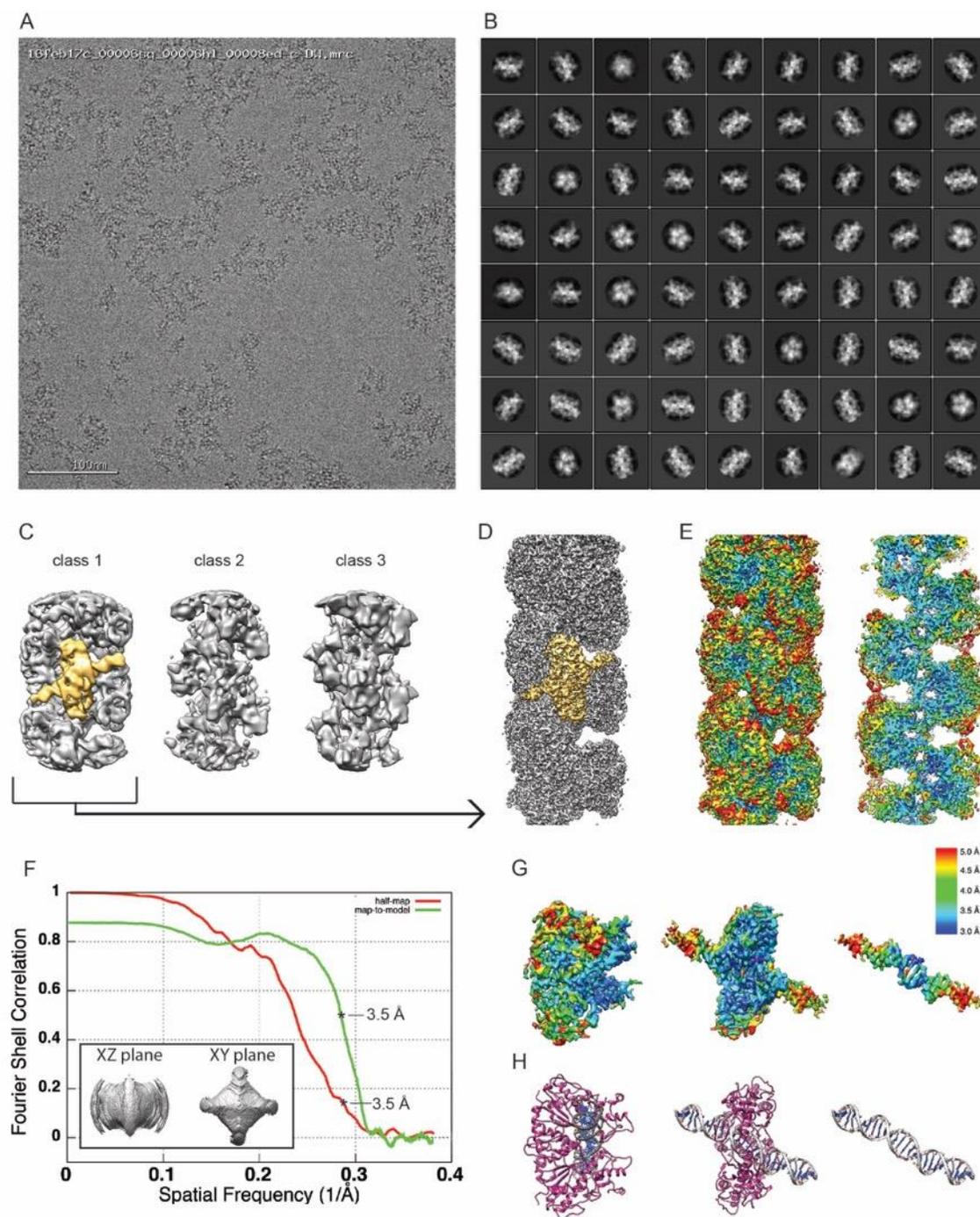


Figure S1. Cryo-EM data, reconstruction and validation. Related to Figure 2. (A) Cryo-EM micrograph of SgrAI. (B) 2D class averages. (C) Reconstructed maps after asymmetric 3D classification showing a class resembling SgrAI (class 1) and two classes with ill-defined features. (D) High-resolution refinement of class 1 from C. In both C and D, a single DBD is highlighted in yellow. (E) Reconstruction from D colored by local

resolution, with the full map at left and a cutaway of the center of the filament at right. (**F**) FSC curves showing half-map resolution (red) and map-to-model resolution (green), both indicating a value of ~ 3.5 Å. 3D-FSC⁴⁹ isosurfaces are displayed at a threshold of 0.5 within the inset for two perpendicular planar views. (**G**) segmented DBD map colored by local resolution with (**H**) the corresponding atomic model.

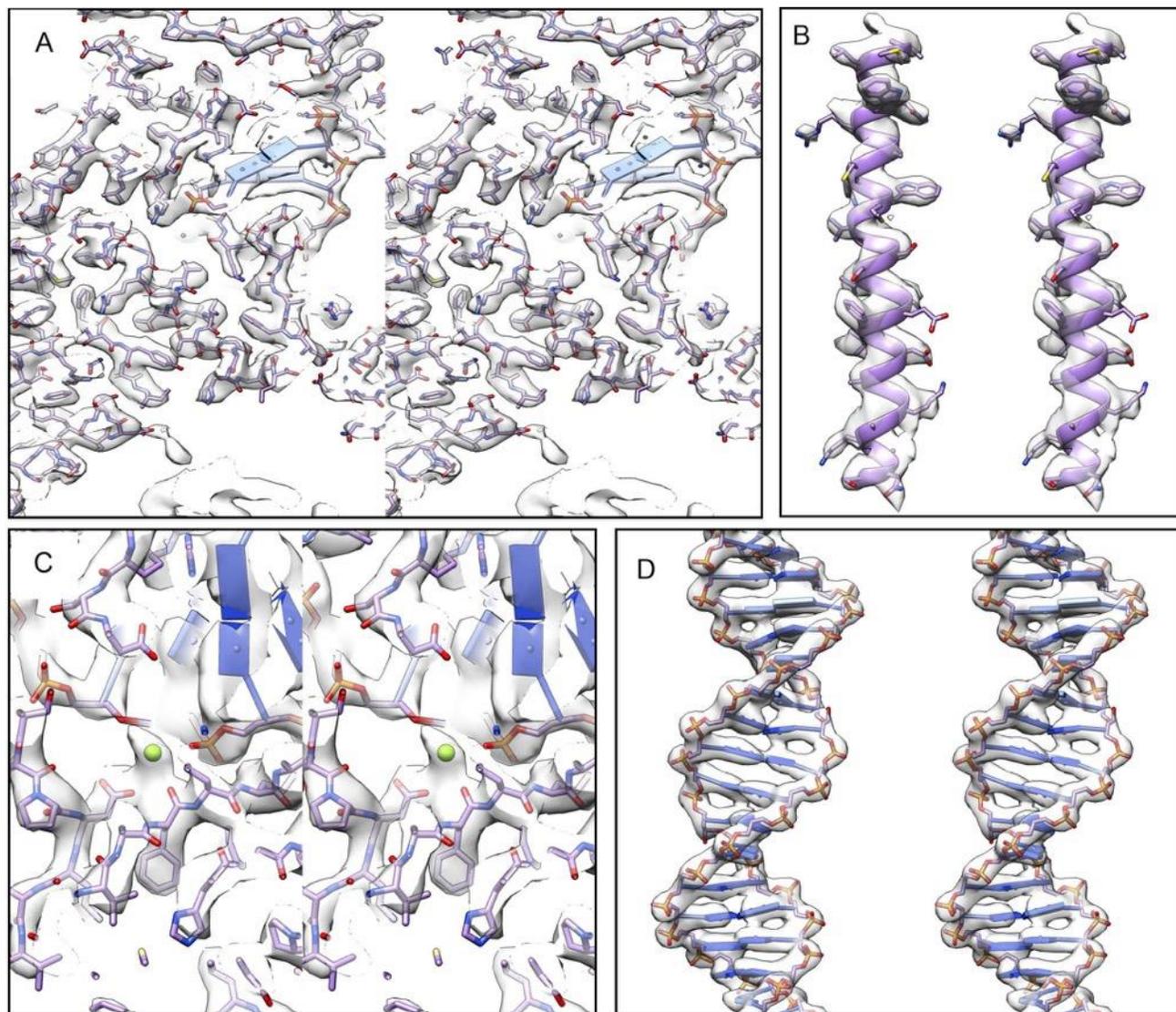


Figure S2. Correspondence between cryo-EM map and atomic model. Related to Figure 1. Stereo views showing (A) a section of the ROO, (B) an alpha helix (residues 90-122), (C) the active site, with bound Mg²⁺ ions (green), and (D) segmented density of the PC DNA.

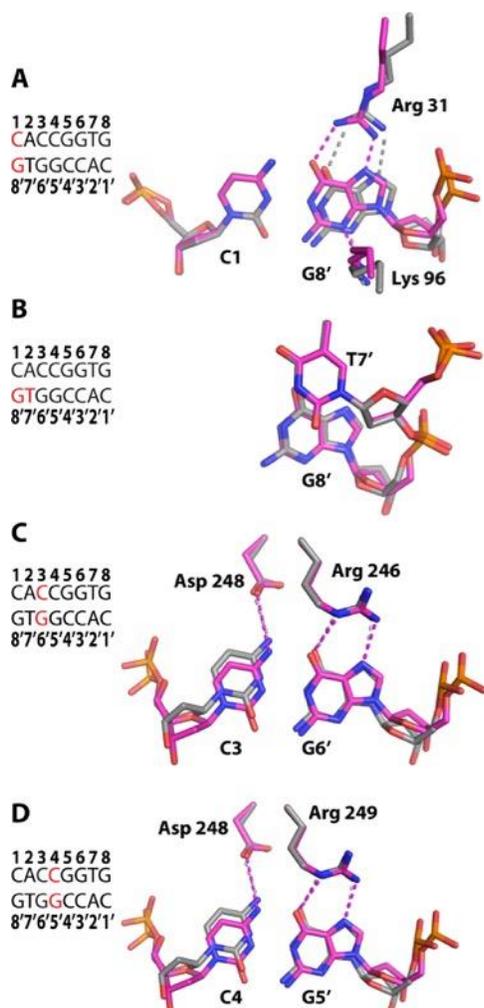


Figure S3. Preservation of contacts between SgrAI and the primary recognition sequence in DNA responsible for sequence specific recognition. Related to Figure 5. (A) The low-activity 3DVO structure shown with carbon atoms in gray overlaid with the high-activity ROO filament cryo-EM structure shown with carbon atoms in magenta. Hydrogen bonds indicated by dashed lines. Lys 96 contacts the N3 of G8', Arg 31 makes hydrogen bonds to the O6 and N7 of G8'. **(B)** Structure at the first base step viewed perpendicular to the base planes. No direct readout contacts are made to the C1 or A2 bases (not shown). Little base stacking (*i.e.* little overlap of the bases in this view) is found between T7' and G8', thought to be part of the indirect readout of the pyrimidine at position 7 of the recognition sequence²⁶. **(C)** As in A, but contacts to the C3/G6' base pair. Arg 246 makes hydrogen bonds to the N7 and O6 of G6'. Asp 248 makes a hydrogen bond to the N4 of C3. **(D)** As in A, but contacts to the C4:G5' bp. Asp 248 makes a hydrogen bond to the N4 of C4, and Arg 249 makes

hydrogen bonds to the N7 and O6 of G5'. In all panels, blue, red, and orange refer to nitrogen, oxygen, and phosphorus, respectively.

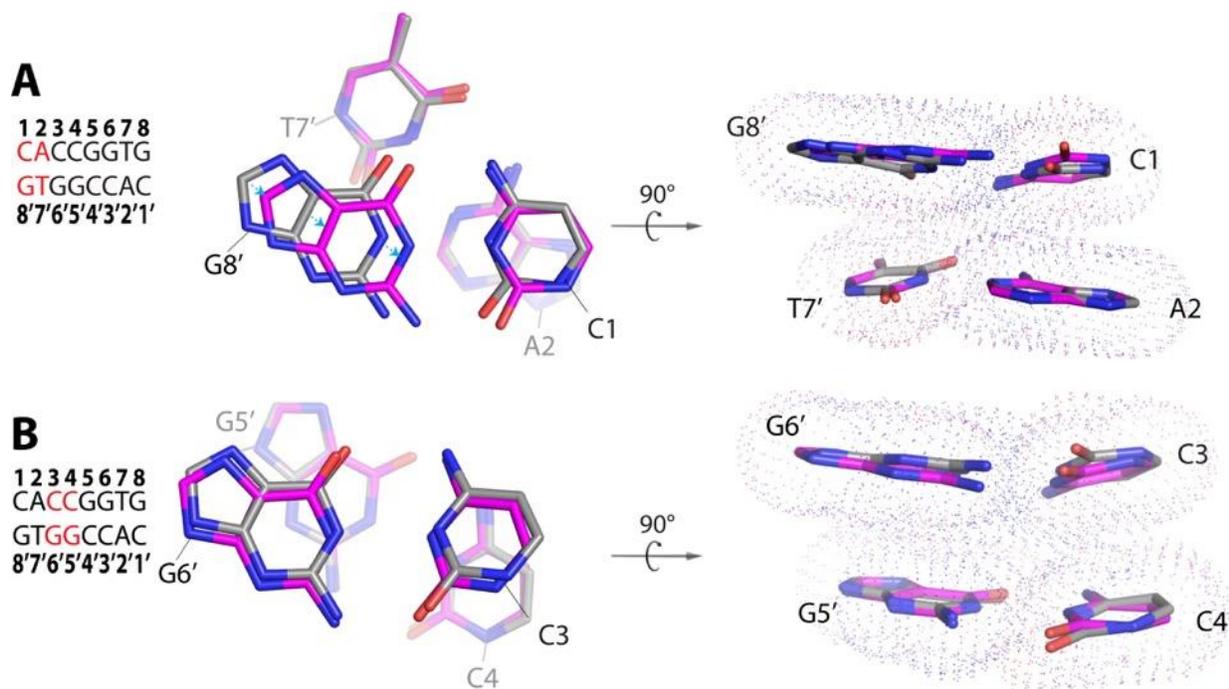


Figure S4. Base stacking comparison between T and R states. Related to Figure 5. (A) Base Stacking at the first base step. All atoms of the base pair in the background were used in the superposition. Atoms from the cryo-EM ROO filament model representing the R state shown with carbon atoms in magenta. Those of 3DVO, representing the T state, shown with carbon atoms in grey. A shift in the position of the G8' base results in a slight increase in stacking overlap with T7'. **(B)** As in A, but the third base step. No significant difference is observed.

