CRITICAL REVIEWS IN BIOCHEMISTRY AND MOLECULAR BIOLOGY
https://doi.org/10.1080/10409238.2024.2440315

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

Structures, mechanisms, and kinetic advantages of the SgrAI filament forming mechanism

Nancy C. Horton[a] and Dmitry Lyumkis[b,c,d]

[a]Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona, USA; [b]The Salk Institute for Biological Studies, La Jolla, California, USA; [c]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, California, USA; [d]Graduate School of Biological Sciences, Section of Molecular Biology, University of California San Diego, La Jolla, California, USA

ABSTRACT

This review documents investigations leading to the unprecedented discovery of filamentation as a mode of enzyme regulation in the type II restriction endonuclease SgrAI. Filamentation is defined here as linear or helical polymerization of a single enzyme as occurs for SgrAI, and has now been shown to occur in many other enzyme systems, including conserved metabolic enzymes. In the case of SgrAI, filamentation activates the DNA cleavage rate by up to 1000-fold and also alters the enzyme's DNA sequence specificity. The investigations began with the observation that SgrAI cleaves two types of recognition sequences, primary and secondary, but cleaves the secondary sequences only when present on the same DNA as at least one primary. DNA cleavage rate measurements showed how the primary sequence is both a substrate and an allosteric effector of SgrAI. Biophysical measurements indicated that the activated form of SgrAI, stimulated by binding to the primary sequence, consisted of varied numbers of the SgrAI bound to DNA. Structural studies revealed the activated state of SgrAI as a left-handed helical filament which stabilizes an altered enzyme conformation, which binds a second divalent cation in the active site. Efforts to determine the mechanism of DNA sequence specificity alteration are ongoing and current models are discussed. Finally, global kinetic modeling of the filament mediated DNA cleavage reaction and simulations of *in vivo* activity suggest that the filament mechanism evolved to rapidly cleave invading DNA while protecting the *Streptomyces* host genome.

ARTICLE HISTORY
Received 27 September 2024
Revised 4 December 2024
Accepted 6 December 2024

KEYWORDS
Enzyme mechanism; enzyme filament; enzyme polymer; restriction endonuclease; allostery; DNA binding protein; DNA nuclease

Introduction

Filament forming enzymes

Polymer or filament formation by enzymes has become increasingly recognized as an important regulatory mechanism in the cell (Park and Horton 2019, 2020; Guo and Liu 2023; Hvorecny and Kollman 2023). Although the formation of linear, helical, or cylindrical filaments (Figure 1(A,B)) by metabolic enzymes has been known for over 50 years, only in the last decade has their biological relevance become appreciated (Park and Horton 2019, 2020). Filamentation is known to change enzymatic characteristics, such as cooperativity, reaction kinetics, substrate specificity, and the affinity for ligands (substrates, products, or allosteric effectors). In some cases, filamentation can even completely alter the type of activity performed by the enzyme (Teixeira et al. 2019). Examples of filament forming enzymes can be found in diverse cell types including bacteria, yeast, plants, and metazoans (Park and Horton 2019). Enzymes in their filamentous states mediate diverse biochemical functions in the cell, including intracellular signaling, transcription regulation, redox regulation, and metabolism (Park and Horton 2019). Many filament forming enzymes are found at key points in metabolic pathways that are also highly regulated by numerous allosteric effectors, and filamentation serves as an additional layer of regulation (Hvorecny and Kollman 2023). In some cases, the filament structure locks the enzyme into a single conformation, which can be active in some cases or inactive in others (Park and Horton 2019). The formation of inactive enzyme filaments may be relevant to help sequester enzymes when they are not needed in the cell, while allowing for rapid redeployment, as necessary. A function of buffering active enzyme

CONTACT Nancy C. Horton nhorton@arizona.edu Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA.

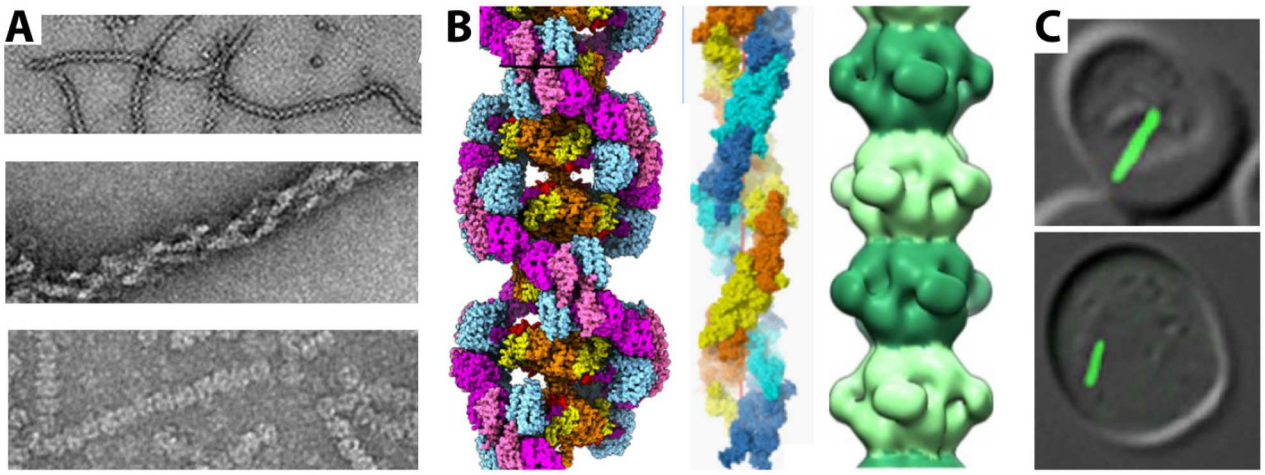

Figure 1. Filaments and cellular self-assemblies (CSA) of various filament forming enzymes. A. TEM images of filaments of Δ1-pyrrole-5-carboxylate synthetase (P5CS, top) (Zhong et al. 2022), glutaminase (center) (Ferreira et al. 2013), and IMPDH (lower) (Anthony et al. 2017). B. Models of enzyme filaments: Δ1-pyrrole-5-carboxylate synthetase (P5CS, left) (Zhong et al. 2022), glutaminase (center) (Ferreira et al. 2013), and IMPDH (right) (Anthony et al. 2017). C. Self-assemblies in yeast cells: glutamate synthase (upper) (Noree et al. 2010), and CTP synthase (lower) (Noree et al. 2010).

concentrations in the cell has also been suggested (Aughey et al. 2014). Filaments that are in an activated state may serve to rapidly and cooperatively upregulate enzyme activity when needed. The modulation of allosteric effector binding affinities allows for different set points for up or down-regulation by those effectors when needed (Lynch et al. 2020; Simonet et al. 2020; Hvorecny and Kollman 2023). Some enzyme filaments are capable of switching between active and inactive conformations, which allows for greater cooperativity and coordination of these states (Lynch and Kollman 2020). Other filamenting enzymes form more than one type of filament, which may be active, inactive, or modulated in other characteristics (Hunkeler et al. 2018). Post-translational modifications or the binding of other proteins can also affect filament stability (Calise et al. 2024).

Many filamenting enzymes form large reversible self-assemblies in cells that can be visualized using fluorescence microscopy (Figure 1(C)) (Narayanaswamy et al. 2009; Werner et al. 2009; Liu 2010; Noree et al. 2010; Ibstedt et al. 2014; Suresh et al. 2015; Shen et al. 2016). Formation of polymeric enzyme filaments *in vitro* has in some cases been linked to cellular self-assembly (CSA) formation. For example, point mutations that disrupt filaments of a mammalian PFK-1 enzyme *in vitro* also disrupt the CSAs (Petrovska et al. 2014; Webb et al. 2017; Keppeke et al. 2018). However, little data exists as to the structure of enzymes within the CSA, though in a small number of cases the CSA has been shown to be composed of bundles of enzyme filaments (Petrovska et al. 2014; Dietrich et al. 2022; Hugener et al. 2024). CSA may also contain more than one type of enzyme and may contain liquid-like properties and even RNA (Fuller et al. 2020). Formation of CSA often occurs during cellular stress but is reversible (i.e. they can be rapidly dispersed under physiological conditions) thereby distinguishing them from other cellular protein assemblies such as amyloid.

SgrAI is not your "classic" type II restriction endonuclease

SgrAI is activated by its primary recognition sequence that also expands its specificity

SgrAI is a type IIF restriction endonuclease from *Streptomyces cyaneofuscatus* (Genbank GCA_035578105.1, previously thought to be from *Streptomyces griseus*) (Tautz et al. 1990). Restriction endonucleases (REs) are key players that mediate innate immunity in bacteria and protect their host from phage infection (Pingoud 2004). In many respects, SgrAI exhibits "typical" type II RE characteristics (i.e. similar to enzymes such as EcoRI, EcoRV, BamHI, PvuII, which have been studied in great detail), such as symmetric Mg^{2+}-dependent cleavage of each strand of dsDNA within a palindromic recognition sequence (Roberts et al. 1993; Aggarwal 1995; Pingoud and Jeltsch 2001; Pingoud 2004; Pingoud et al. 2005). However, SgrAI also exhibits more unusual activities, such as the requirement of a longer recognition sequence (8 bp rather than 4-6 bp) and faster DNA cleavage in the presence of multiple copies of its recognition sequence within the same contiguous DNA. The presence of these unusual activities therefore required placing SgrAI in the type IIF RE category, along with other enzymes such as SfiI (Wentzell et al. 1995), Cfr10I

(Siksnys et al. 1999), Bse634I (Grazulis et al. 2002), and NgoMIV (Deibert et al. 2000). Significantly, SgrAI has additional unique properties that are not encountered in any other RE known to date. It is the only RE that exhibits an unusual secondary DNA sequence specificity, which can occur under physiological reaction conditions but requires the presence of its primary (i.e. cognate) sequence. SgrAI is also the only RE known to form filaments as part of its DNA cleavage pathway.

The cognate DNA recognition sequence of SgrAI was initially identified as the three distinct sequences with the pattern CR|CCGGYG, where R=A or G and Y=C or T (Tautz et al. 1990). Cleavage in the phosphodiester bond occurs after the R at the second nucleotide position, leaving 5′CCGG single stranded "sticky" ends on the cleaved DNA. Because of their relatively longer recognition sites (8 bp compared to earlier characterized enzymes that have 4-6 bp recognition sites) "rare cutting" restriction endonucleases such as SgrAI have historically been sought for sequencing and related applications (Lambert et al. 2008; Tümmler 2014). However, early studies showed that SgrAI will cleave plasmid DNA up to 30-fold faster when more than one primary recognition site is present compared to plasmids with a single site (Figure 2(A)) (Bilcock et al. 1999). Additional studies indicated that 14 different DNA sequences that differ from the primary sequence (denoted secondary) were cleaved if present on a plasmid with at least one primary sequence, but not in its absence (Figure 2(B)) (Bitinaite and Schildkraut 2002). This effect of the primary sequence on the rate at which SgrAI cleaves DNA containing primary recognition sites is recapitulated *in trans* using an oligonucleotide containing ends mimicking the cleaved primary sequence (Figure 2(C)) (Bitinaite and Schildkraut 2002). This experiment also showed that the cleaved primary sequence is an activator of SgrAI.

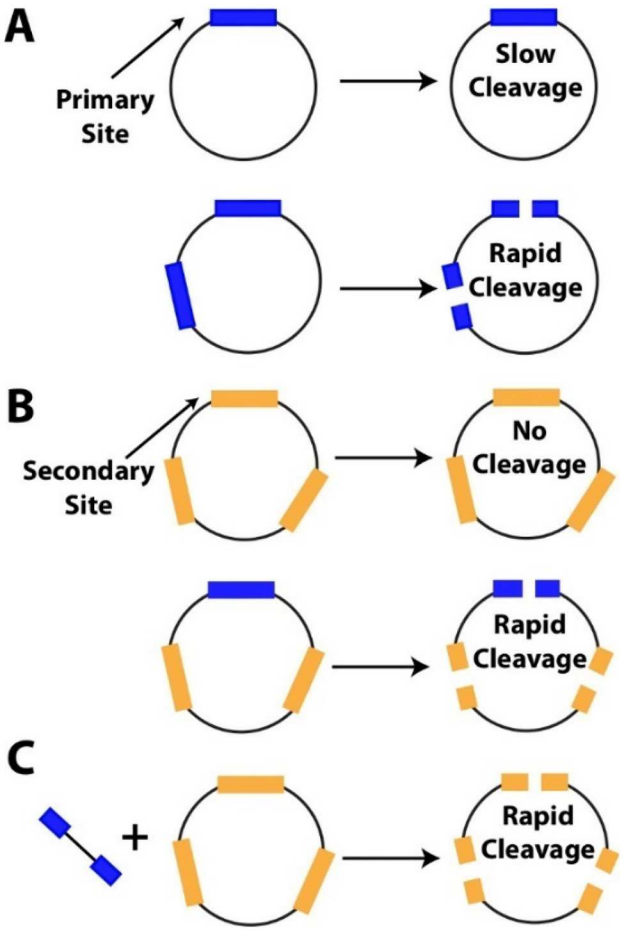

Figure 2. Unusual DNA cleavage activity of SgrAI. A. (Upper) Plasmid DNA with a single primary recognition site is cleaved 30-fold slower than (lower) a plasmid with two copies (Bilcock et al. 1999). B. (Upper) a plasmid with three secondary sequences is not cleaved at detectable levels even after 20 h (Bitinaite and Schildkraut 2002), however (lower) substitution of one secondary with a primary sequence results in rapid cleavage of all three sites (Bitinaite and Schildkraut 2002). C. This effect works in trans as well as in cis. Addition an oligonucleotide with end containing the cleaved primary sequence induces rapid cleavage of a plasmid with three secondary sites and no primary sites (Bitinaite and Schildkraut 2002).

Evidence of communication between SgrAI bound to separate recognition sequences

Secondary sequences (or sites) that are cleaved by SgrAI contain the patterns **D**RCCGGYG and C**C**CCGGYG, wherein the bold underlined nucleotides are the sequence substitutions. SgrAI cleaves secondary sites, under most circumstances, only after initial binding and recognition of a cognate primary site DNA. The lack of cleavage of the secondary sequences when there are no primary sequences present (Bitinaite and Schildkraut 2002) cannot be simply attributed to a lack of DNA binding, since SgrAI binds to secondary sequences nearly as tightly as it binds to primary sequences, with K_Ds in the low nM range (Daniels et al. 2003; Hingorani-Varma and Bitinaite 2003). Further studies also suggest the formation of higher-order complexes by DNA bound SgrAI. For example, single molecule experiments using optical tweezers provided evidence for DNA looping by SgrAI on long DNA molecules containing multiple sites (Gemmen et al., 2006, 2006). In addition, the steady state cleavage rate constant of a plasmid with two primary sites, but not with one primary site, increased with increasing concentrations of SgrAI protein. This suggests cooperativity that is dependent upon SgrAI concentration, if and only if there are opportunities for bridging two separate recognition sequences (Daniels et al. 2003). A full steady-state kinetic analysis showed high levels of

cooperativity on the concentration of DNA substrates (both primary and secondary sequences), with Hill coefficients between 2 and 3 (Hingorani-Varma and Bitinaite 2003). This was true for both the primary and secondary recognition sequences. The addition of the stimulatory oligonucleotide with ends containing the cleaved primary sequence increased the rate of DNA cleavage by 3-4 times for primary sequences, but up to 100 times faster for secondary sequences. Furthermore, there was an observed increase in the specificity factor k_{cat}/K_M for both types of sequences (*via* effects on k_{cat} in the case of primary sequences, but effects on both k_{cat} and K_M for secondary). The added oligonucleotide also reduced the Hill coefficient to ~1, indicating that the cooperativity on substrate concentration was lost (Hingorani-Varma and Bitinaite 2003). Hence, the presence of the activator DNA, which contains the cleaved primary sequence, eliminates the cooperativity on the DNA substrate concentration, presumably by acting in its place (Hingorani-Varma and Bitinaite 2003).

Other Type IIF REs were shown to form tetramers with 2 DNA binding sites. However, sedimentation velocity and sedimentation equilibrium analysis with an analytical ultracentrifuge showed that SgrAI is dimeric in the absence of bound DNA (Daniels et al. 2003). Because the primary recognition sequence can act as an activator as well as a substrate, the presence of at least 2 DNA binding sites on the activated SgrAI molecule is implicated, however the sedimentation studies showed a 1:1 binding stoichiometry of the SgrAI dimer for duplex DNA (whether primary or secondary sequence) (Daniels et al. 2003). To obtain 2 DNA binding sites, assembly of two DNA bound SgrAI dimers (DBD) into a functional tetramer could be possible, yet the sedimentation studies did not identify this species either. Instead, a species that is much larger than 2 DBD was observed (Daniels et al. 2003). This high molecular weight species (HWMS) only formed with the addition of primary sequences, but not with the addition of secondary sequences. However this early observation was dismissed as an irrelevant aggregation of SgrAI in its apo form onto SgrAI bound to DNA, and that the elusive tetrameric SgrAI must be a short-lived species (Daniels et al. 2003).

Additional biochemical and single molecule studies also pointed to the necessity of assembling at least two DBDs to form the activated SgrAI (Wood et al. 2005). A study of SgrAI cleavage on linked plasmids (connected by concatenation) showed that cleavage of two primary sites on the same contiguous DNA proceeds much faster than cleavage of two primary sites on separate DNA molecules (Wood et al. 2005). This indicated that the communication between recognition sites occurs through three-dimensional space and not through a one-dimensional sliding mechanism thought to occur among some REs (Szczelkun 2002). The cleavage of a secondary sequence on a plasmid concatenated with another plasmid that contains a primary sequence proceeded faster in comparison to the case when the secondary and primary sites reside on separate (unconcatenated) plasmids. However, when the two sites reside on separate but concatenated plasmids, rather than on a single contiguous plasmid, the reaction does not go to completion. This is likely because the cleavage of the primary sequence on one plasmid results in the loss of the concatenation, leaving the plasmid containing the secondary sequence disconnected from the plasmid containing the primary sequence (Wood et al. 2005). Using oligonucleotides, a requirement for longer flanking bp (the bp on either side of the 8 bp recognition sequence) was found to be important, as a primary recognition sequence embedded in a 30 bp, but not 20 bp oligonucleotide, stimulates cleavage *in trans* of primary or secondary sequences which were also embedded in 30 bp (Wood et al. 2005).

X-ray crystallographic structures of SgrAI bound to DNA shows a DNA bound dimer (DBD)

The early biochemical studies suggested the formation of a higher order structure or assembly that enhances the intrinsic DNA cleavage rate constant of SgrAI. However, crystal structures of SgrAI bound to its primary recognition sequence embedded in an 18 bp double-stranded oligonucleotide showed only the SgrAI homodimer bound to a single DNA duplex (Figure 3(A)) (Dunten et al. 2008). The mode of DNA binding by SgrAI is similar to that of other type II REs, and especially NgoMIV, which shares the inner 6 bp of the SgrAI primary recognition sequence (Deibert et al. 1999, 2000). To trap the complex prior to DNA cleavage, Ca^{2+} was substituted for the biologically relevant cofactor Mg^{2+}, and two Ca^{2+} were observed in each active site (Figure 3(B)). The first Ca^{2+} is located in the canonical site A of the "Two-Metal Ion Mechanism", as described for many divalent metal ion-dependent DNA nucleases (and other phosphoryl transfer enzymes) (Figure 3(C)) (Beese and Steitz 1991; Kim and Wyckoff 1991; Steitz and Steitz 1993; Sträter et al. 1996; Horton et al. 2002a; Yang et al. 2006; Horton 2008; Yang 2011). In this mechanism, two divalent cations, typically Mg^{2+}, bind near the scissile phosphate (SP, i.e. the phosphodiester bond to be cleaved) and perform various critical functions to catalyze the cleavage reaction such as

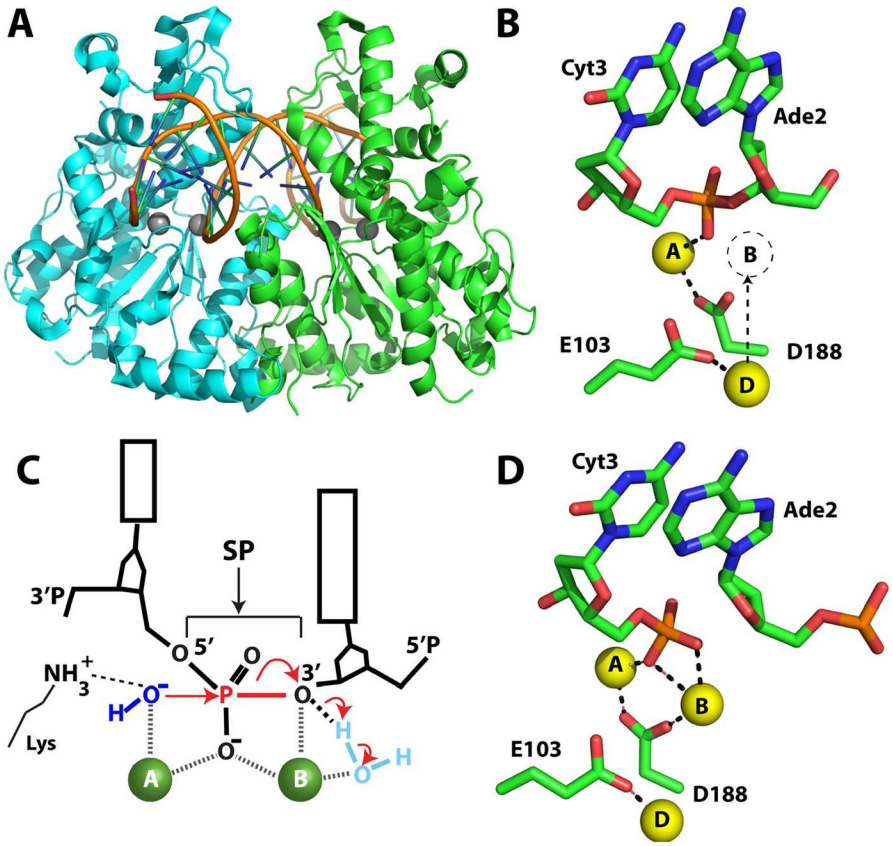

Figure 3. Non-filamentous SgrAI bound to DNA and the divalent cation positions in the active site. A. Ribbon diagram of the SgrAI homodimer shown in cyan and green bound to an 18 bp duplex containing the 8 bp primary sequence CACCGGTG. The two Ca^{2+} ions bound in each active site are shown as grey spheres (PDB code 3DVO (Dunten et al. 2008)). B. Position of Ca^{2+} ions (yellow spheres) in the active site of SgrAI (based on PDB code 3DVO (Dunten et al. 2008)). C. Idealized two metal ion mechanism proposed for many DNA nucleases (Beese and Steitz 1991; Kim and Wyckoff 1991; Steitz and Steitz 1993; Sträter et al. 1996; Horton et al. 2002a; Yang et al. 2006; Horton 2008; Yang 2011). D. Position of Mg^{2+} ions in the crystal structure with cleaved DNA (PDB code 3MQY (Little et al. 2011)).

positioning and activation of the nucleophile (water or hydroxide) for nucleophilic attack on the phosphorus atom of the SP, as well as positioning and stabilizing the SP and the O3′ leaving group. In the case of the SgrAI structure, the second divalent cation position in site B was unoccupied. The second Ca^{2+} ion was instead located in a distal site, site D (Figure 3(B)). We wondered if this mispositioning might be due to the fact that Ca^{2+} cannot be used by SgrAI to catalyze DNA cleavage, and therefore we tested the binding of Mn^{2+}, which can also confer DNA cleavage activity on SgrAI. However, only sites A and D were again seen to be occupied, and the DNA was found in the uncleaved state (Dunten et al. 2008). Several lines of evidence suggest that this structure is that of the low activity conformation of SgrAI. First, SgrAI forms the same dimeric structure, with the same positions of metal ions, when bound to a secondary sequence (Little et al. 2011), but secondary sequences are cleaved very slowly (Bitinaite and Schildkraut 2002; Hingorani-Varma and Bitinaite 2003). Second, plasmids containing only a single primary recognition sequence are also cleaved very slowly by SgrAI (Bilcock et al. 1999); in this case, SgrAI is presumably a dimer bound in a 1:1 complex with its recognition sequence, i.e. it has not had a chance to oligomerize (Daniels et al. 2003). Third, the apparent mispositioning of Ca^{2+} and the catalytically competent Mn^{2+} in site D, with site B unoccupied, also suggests that this dimeric structure represents the low activity state. Therefore, we proposed that the crystal structures of dimeric DNA bound SgrAI is of its low DNA cleavage activity state, and that activation involves a conformational change resulting in a shift of the site D ion to a newly formed site B (and/or occupation of site B in addition to site D) (dashed arrow, Figure 3(B)) (Dunten et al. 2008).

Next, we sought to determine the structure of activated SgrAI. Because activation of SgrAI was shown to occur with the cleaved version of the primary sequence (Bitinaite and Schildkraut 2002), the structure of SgrAI

bound to cleaved DNA was determined (Little et al. 2011). However, the structure was again found to consist of dimeric SgrAI bound to a single duplex of DNA, only in this case the DNA was cleaved. The structure was otherwise unchanged from those previously determined, with the exception of a small shift in the bound DNA to accommodate the increased distance between atoms at the cleaved bond, and the occupation of 3 metal ion sites (sites A, B, and D) (Figure 3(D)). To explain the occupation of Mg^{2+} in site B, we proposed that the cleaved phosphate moiety stabilizes binding of the site B Mg^{2+}.

Activation by simultaneous binding to two copies of its recognition sequence had been well established for the related type IIF enzymes such as SfiI, NgoMIV, Cfr10I, and Bse634I (Wentzell et al. 1995; Siksnys et al. 1999; Deibert et al. 2000Manakova et al., 2012; Grazulis et al. 2002; Vanamee et al. 2005). However, these enzymes form stable tetramers (i.e. two DBD), and are ***not*** known to have expanded DNA sequence specificity, as seen in the secondary activity exhibited by SgrAI. Because the recognition sequence and protein structural fold are most similar between SgrAI (primary recognition sequence: CR|CCGGYG) and NgoMIV (recognition sequence: G|CCGGC) and Bse634I (recognition sequence: R|CCGGY), their tetramer structures were compared to the dimeric structure of SgrAI bound to DNA. The tetramers formed by NgoMIV and Bse634I both assemble as "back-to-back" dimers of dimers with their DNA binding clefts facing away from the tetramer interface (Figure 4(A,B)), however the details of their dimer-dimer interfaces vary considerably (Deibert et al. 2000; Manakova et al. 2012). In the case of NgoMIV, a loop extends from each subunit to contact a subunit of the opposing dimer (spheres and arrow, Figure 4(A)), which is not present in Bse634I or SgrAI. The Bse634I dimers are held together by contacts involving the extension of an alpha helix in each subunit (arrow, Figure 4(B)), which is not present in NgoMIV or SgrAI. When the SgrAI dimer is superimposed onto the NgoMIV tetramer, many steric collisions at the interface result (blue box, Figure 4(C)). On the other hand, when the Bse634I tetramer is used, without the extended alpha-helices, no contacts are made between dimers (blue box, Figure 4(D)). In addition, SgrAI contains segments that are not found in NgoMIV or Bse634I, which occur at the opposite face of the dimer, near the DNA binding cleft (red ovals, Figure 4(C,D)). These were subsequently found to be important in filament formation by SgrAI (to be described below). In summary, none of these models could explain the structural basis for SgrAI activation, at least based on the known RE homologs at the time.

SgrAI forms an activated high molecular weight species (HMWS)

Formation of the HMWS is dependent on the concentration of SgrAI bound to a primary site sequence. To find conditions that favor the activated state of SgrAI bound to DNA, to allow for capturing the activated state using structural methods, a series of biophysical studies were undertaken (Park et al. 2010). First, native gel electrophoresis was used with oligonucleotides containing primary sequences either embedded into a short construct (termed 18-1 with 18 bp of duplex DNA embedding the primary site) or embedded into a long construct (termed 40-1 with 40 bp of duplex DNA embedding the primary site). The long construct also had three variations of the DNA in the cleavage site: the DNA was either uncleaved (40-1), pre-cleaved (PCP), or pre-cleaved but missing the phosphate at the cleavage site (PC). The rationale was to identify the tetramer in native gel electrophoresis by the expected pattern of 3 shifted bands derived from a tetrameric SgrAI bound to 1) two short DNAs, 2) two long DNAs, and 3) one short and one long DNA. To mark the position of the DNA bound SgrAI dimer (DBD) in native PAGE, SgrAI bound to a DNA duplex containing a secondary site was used (lanes 7-9, Figure 5(A)). However, this pattern of three super-shifted bands (in addition to that for the DBD) was never seen. Instead, only a single high molecular weight species (HMWS) was observed in addition to the DBD, but only when both the concentrations of SgrAI and DNA are high enough. For example, SgrAI bound to 18-1 formed only DBD with low concentrations of SgrAI and DNA (1 nM and 50 nM, respectively, lane 2, Figure 5(A)), but formed DBD and HMWS when concentrations of both 18-1 and SgrAI were high (1 µM each, lane 4, Figure 5(A)). High concentrations of SgrAI (1 µM) with low DNA concentration (1 nM ^{32}P-18-1) does not give rise to HMWS (lane 3, Figure 5(A)). In addition to the requirement for relatively high concentrations of both SgrAI and DNA, the presence of a DNA containing the primary sequence is required. Lanes 7-10 of Figure 5(A) show that the HMWS is not formed when a single base pair substitution is introduced into the 18-1 DNA construct to create a secondary sequence. Only the DBD is formed regardless of SgrAI or DNA concentration.

The oligonucleotide containing longer flanking base pairs and a cleaved primary site (i.e. PCP, a pre-cleaved primary sequence embedded in a 40 bp DNA with 16 bp flanking bp on either side of the recognition site) showed the formation of the HMWS under similar conditions as 18-1 (i.e. high concentrations of DNA and SgrAI) (lane 14, Figure 5(A)) (Park et al. 2010). A minor species was also observed, which ran at an intermediate length between the DBD and the HMWS (marked with a *, lane 13, Figure 5(A)). This minor species was only found with 1 nM PCP and 1 µM SgrAI. The species could be a DBD bound to annealed PCP that migrates

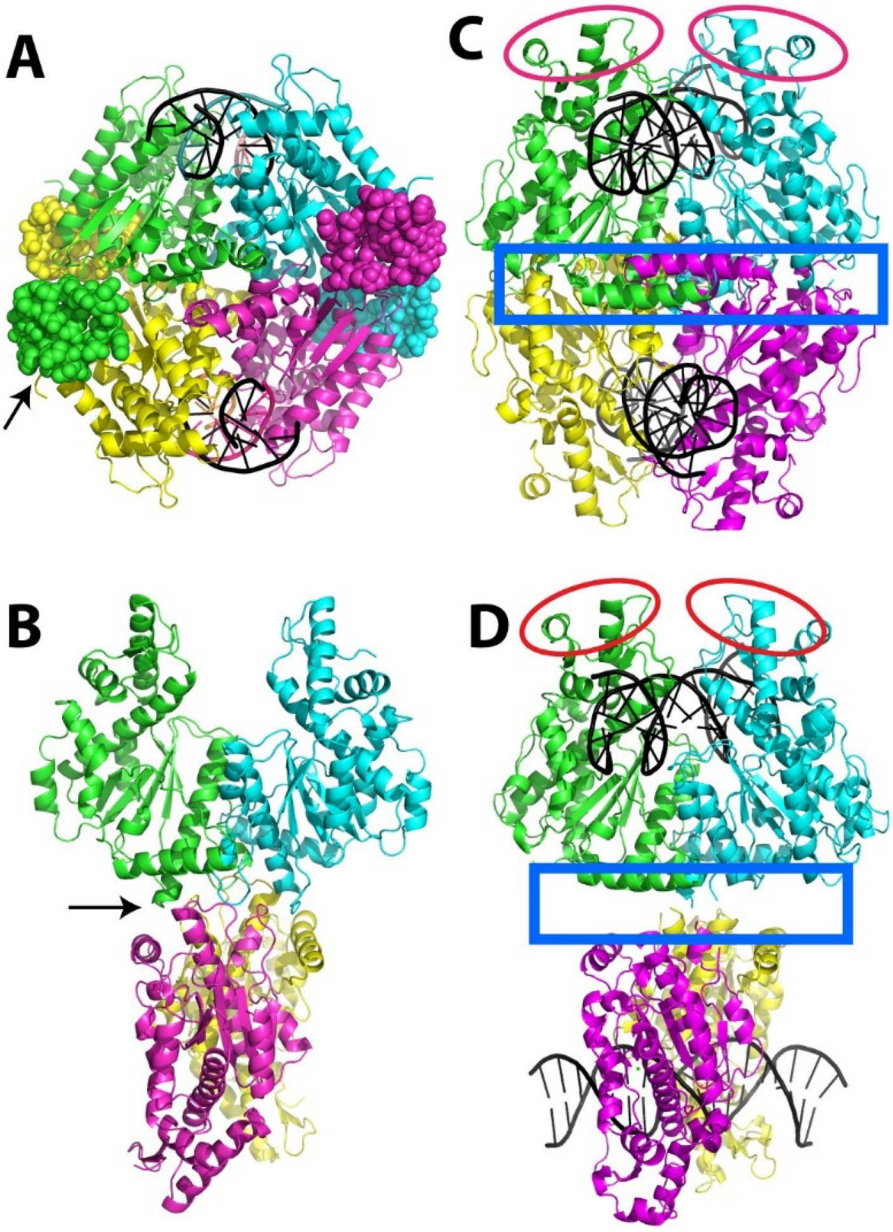

Figure 4. SgrAI Tetramer models based on NgoMIV and Bse634I tetramer structures. A. Tetrameric structure of DNA bound NgoMIV (PDB code 1FIU (Deibert et al. 2000)) colored by chain. Spheres and arrow identify a loop which forms a large interface between subunits of opposing dimers. B. Tetrameric structure of DNA bound Bse634I (PDB code 3V21 (Manakova et al. 2012)) colored by chain. The arrow indicates helical extensions absent in SgrAI but important for the dimer-dimer interface in the Bse634I tetramer. C. Model of SgrAI tetramer prepared by aligning each DNA bound SgrAI dimer onto each of the two dimers in the NgoMIV tetramer. Steric clashes at the dimer-dimer interface are outlined by a blue box. Red ovals outline inserts in the SgrAI structure found to be important in filament formation. D. Tetrameric model of SgrAI based on Bse634I where the DNA bound SgrAI dimer was aligned with each dimer in the Bse634I tetramer. The blue box outlines missing contacts at the dimer-dimer interface. Red ovals outline inserts in the SgrAI structure found to be important in filament formation.

more slowly due to the longer flanking bp. Alternatively, it could be a species containing two or more DBDs or a DBD with an additional copy of SgrAI bound. Finally, we tested the effect of adding this longer DNA, PCP, to reactions with ^{32}P-18-1 or ^{32}P-18-2 (identical to 18-1, but with a secondary sequence substitution, C**C**CCGGTG) and in the presence of high concentrations of SgrAI (lanes 16 and 18, Figure 5(A), respectively). The PCP unsurprisingly induced HMWS formation of SgrAI bound to ^{32}P-18-1, but also with ^{32}P-18-2. Thus, while the DNA with a secondary sequence (18-2) will not induce HMWS even with excess DNA and SgrAI (conditions that induce HMWS with 18-1), it will join HMWS stimulated by the presence of high concentrations of

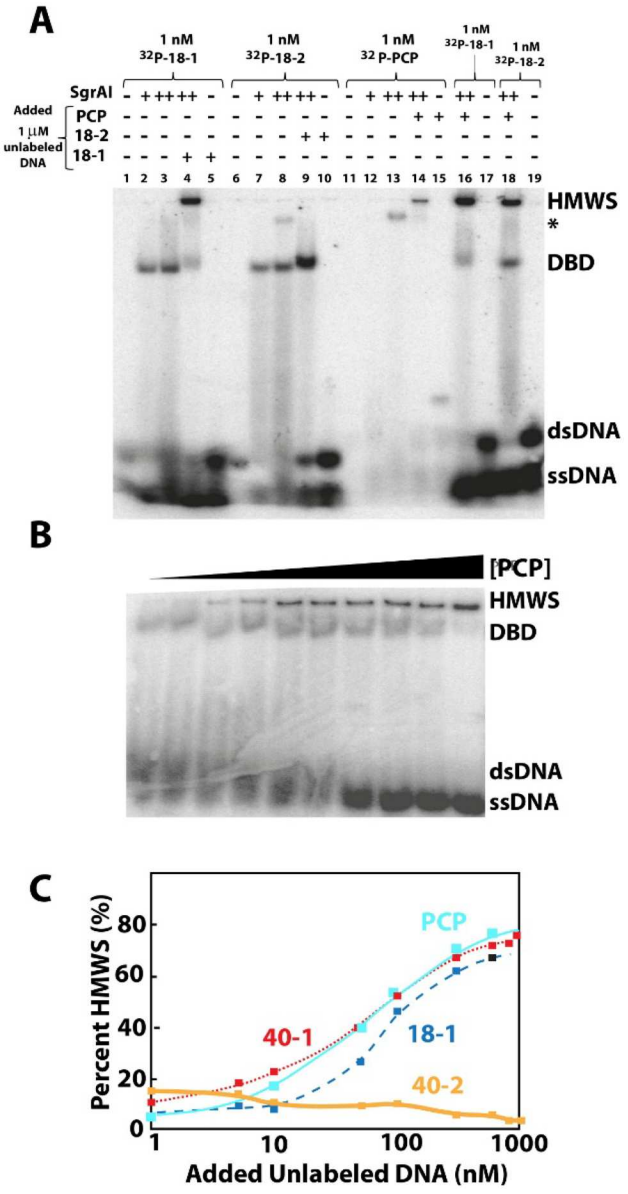

Figure 5. Native gel electrophoresis and analysis. A. Gel shift experiment with varied DNA sequence and SgrAI concentrations. HMWS, high molecular weight species, DBD, DNA bound SgrAI dimer, dsDNA, double-stranded DNA, ssDNA, single stranded DNA. An unknown species likely to be two DBD is marked with a *. B. Titration experiment using native gel electrophoresis showing the increase in HMWS with increasing concentrations of PCP, pre-cleaved 40 bp DNA containing a primary site sequence in the presence of an excess concentration of SgrAI. C. Plot of titration data in terms of the percentage of shifted species (i.e. DBD + HMWS) in the HMWS vs. the concentration of added unlabeled DNA (PCP, 40-1, 18-1, or 40-2). (Adapted from Park et al. (Park et al. 2010)).

SgrAI and PCP. We investigated this effect further by titrating increasing concentrations of PCP in reactions with 1 μM SgrAI and 1 nM ^{32}P-18-1 or ^{32}P-18-2 DNA (Figure 5(B)). Similar results were observed in both cases, where the band intensity of the HMWS increased with increasing concentrations of PCP, and with a corresponding decrease in the DBD band. Figure 5(C) plots the result of several such titrations using 1 nM ^{32}P-18-1, 1 μM SgrAI, and increasing concentrations of PCP, 40-1 (the uncleaved version of PCP), and 40-2 (as in 40-1 however with a single base pair substitution resulting in the secondary sequence C**C**CCGGTG). In all cases where the primary sequence was employed (PCP, 40-1, 18-1, red, dark blue, and light blue, Figure 5(C)), the HMWS was induced with increasing DNA concentrations. In contrast, 40-2, which contained only a secondary sequence, did not induce the HMWS (yellow, Figure 5(C)). Those constructs with the longer flanking bp (PCP and 40-1 with 16 flanking bp) induced HMWS at lower concentrations than 18-1 (with 5 flanking bp) (Park et al. 2010). This result was reminiscent of the earlier report that showed an effect of flanking bp on the activation of DNA cleavage (Daniels et al. 2003). The formation of a large species in our native gel analysis also mirrors what was found in the earlier sedimentation studies, which was dismissed then as an aggregate of excess SgrAI dimers on the DBD (Wood et al. 2005). However, in contrast, we found that high concentrations of SgrAI will not induce HMWS unless the DNA concentration is also high (compare lanes 3-4, Figure 5(A)). Further, we see a consistent titration of HMWS with increasing concentrations of the activating DNA and in the presence of high concentrations of SgrAI (Figure 5(B,C)). Our observations are more consistent with the HMWS being composed of an assembly of the DNA bound form of SgrAI (i.e. the DBD) (Park et al. 2010), rather than of excess SgrAI dimers aggregating on DBD, as suggested earlier (Daniels et al. 2003). In summary, the native gel analysis showed that the HMWS forms only with SgrAI bound to DNA containing a primary sequence, and not to DNA containing a secondary sequence (in the absence of DNA with the primary sequence). Intriguingly, the difference between the primary and secondary DNAs is only a single bp substitution in the recognition sequence, highlighting how small changes to the recognition sequence can have drastic effects on enzyme behavior. The data also show that concentrations of both SgrAI and the primary site containing DNA must be sufficiently high (we found a $K_{1/2}$ around 70 nM in the case of the longer oligonucleotides), and that an oligonucleotide containing the secondary sequence could be drawn into the HWMS if there are sufficient concentration of SgrAI protein bound to the primary sequence(s).

The HMWS is the activated form of SgrAI. As the conditions that lead to the formation of the HMWS became better established, the priority shifted to

assaying enzymatic function. DNA cleavage reactions were used to test the hypothesis that the HMWS is the activated species. Reactions were performed under single turnover conditions with excess (1 µM) SgrAI over the reporter DNA substrate (1 nM ^{32}P labeled DNA). Since the HMWS formed with increasing concentrations of SgrAI bound to a pre-cleaved 40 bp DNA containing a primary recognition sequence (PCP), reactions were measured with increasing concentrations of PCP or PC DNA (which is PCP missing the phosphate at the cleavage site) to mimic the increasing degree of HMWS formed in native gels (Park et al. 2010). Both PCP and PC DNA bind and activate SgrAI indistinguishably (Park et al. 2010). Reactions were quenched with a solution containing 4 M urea and 50 mM EDTA, and then the percentage of cleaved DNA measured by denaturing PAGE and autoradiography. As a result, all steps up to and including DNA cleavage are measured in this method. The resulting data fit well to a single exponential function, indicating that there was a single dominant slow step in the reaction pathway (Park et al. 2010). The cleavage reactions were performed at 37 °C (except where noted), differing from the native gel electrophoresis experiments, which were performed at 4 °C. The reactions also necessitated the use of Mg^{2+} rather than the Ca^{2+} used in the native gel electrophoresis experiments.

The results of the enzymatic cleavage experiments are shown in Figure 6 (note the log plot on the y-axis to facilitate visualization of rate constants different in orders of magnitude). First, in Figure 6(A), the cleavage of ^{32}P labeled 18-1 and 40-1 is shown (Park et al. 2010). Without added activator DNA ("None", 1st column, Figure 6(A)), no HMWS is expected since the concentration of SgrAI bound to DNA is only 1 nM, and a very slow rate constant of ~0.1 min^{-1} (or 9×10^{-4} s^{-1}) is measured. Type II REs typically cleave DNA much faster, for example, EcoRV and EcoRI have DNA cleavage rate constants of 0.6 s^{-1} and 0.3 s^{-1} (36 min^{-1} and 20 min^{-1}), respectively (Lesser et al. 1990; Sam and Perona 1999). The addition of high (0.9 µM) concentrations of 18-1 did not change the measured rate constant for DNA cleavage (Figure 6(A)), even though HMWS were induced at 4 °C in native gels with these conditions. However, consistent with the lack of HMWS in the native gels, a secondary site embedded within a 40 bp construct (40-2) also did not activate DNA cleavage of the 18-1 reporter DNA (Figure 6(A)**)**. Clear activation was however observed with the addition of primary sequences embedded into DNA that contained longer flanking base pairs, such as 0.9 µM 40-1 and 1 µM PCP (Figure 6(A)). Measurements were also performed using ^{32}P labeled 40-1 which showed no activation of DNA cleavage at 1 nM concentration of 40-1, but robust activation at 0.9 µM (Figure 6(A)). Since the same conditions that gave rise to HWMS in the native PAGE studies also led to SgrAI-mediated cleavage activation of DNA containing the primary sequence (40-1, PCP) but not secondary sequence (40-2), we concluded that cleavage activation is dependent upon the concentration of DNA containing the primary sequence. However, the results with 18-1 were not consistent between the native PAGE experiments and the DNA cleavage measurements. HMWS formed with the addition of 0.9 µM unlabeled 18-1 to 1 nM ^{32}P-18-1 and 1 µM SgrAI in the native PAGE experiments performed at 4 °C and with

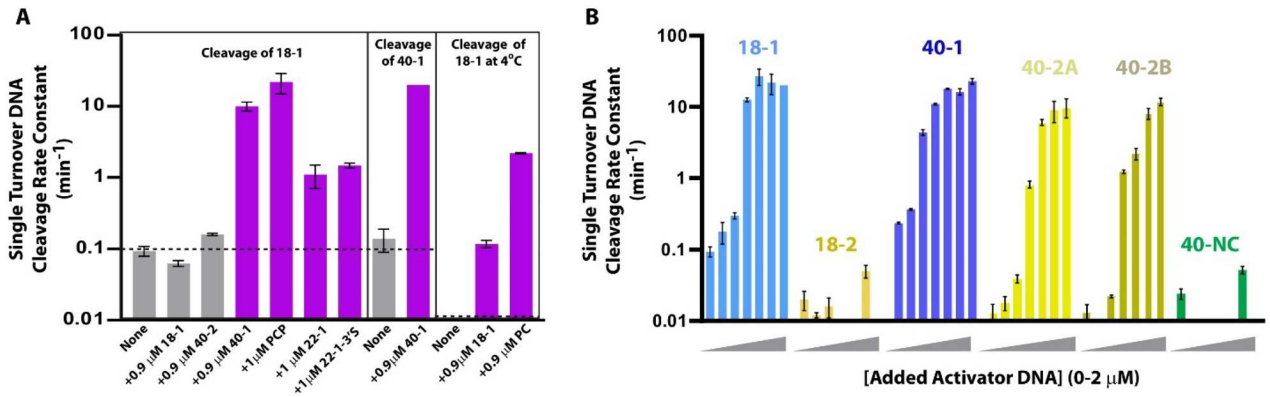

Figure 6. Single turnover DNA cleavage rate constants and the effect of added DNA with primary or secondary sequences in 18 or 40 bp constructs. A. Single turnover DNA cleavage rate constants of ^{32}P labeled 18-1 or 40-1 with or without added unlabeled DNA. 40-2, 40 bp DNA containing a secondary sequence, 40-1, 40 bp DNA containing a primary site, PCP, pre-cleaved 40-1, 22-1, 22 bp DNA containing a primary site, 22-1-3'S, 22-1 with a 3'S substitution at the cleavage site. Grey indicates the absence of significant activation. Purple indicates the activation of DNA cleavage. Reactions performed at 37 °C except where indicated. The dotted line indicates the basal, unactivated DNA cleavage rate constant at 37 °C (Park et al. 2010). B. Single turnover DNA cleavage rate constants at 37 °C of indicated DNA constructs with varied concentrations of added activator DNA (PCP or PC DNA). 40-NC, 40 bp DNA with a noncognate sequence. (Park et al. 2010; Shah et al. 2015).

Ca^{2+} (Figure 5(A)), but activation did not occur in cleavage reactions performed at 37 °C with Mg^{2+} but with otherwise identical conditions (Park et al. 2010). To investigate further, cleavage reactions were performed at 4 °C. The cleavage of 1 nM ^{32}P-18-1 with 1 µM SgrAI at 4 °C was not detectable after 24 h, and therefore the rate constant is estimated as $\leq 4\times10^{-5}\,min^{-1}$ ("None" third column from right, Figure 6(A)). However, with the addition of 0.9 µM unlabeled 18-1, a rate constant of $0.118\pm0.014\,min^{-1}$ was measured showing a ~3000-fold increase. PC DNA at 0.9 µM showed an even greater increase to $2.2\pm0.04\,min^{-1}$, a ~60,000-fold increase. Hence, just as in prior investigations, we concluded that both the length of the flanking base pairs as well as the DNA sequence (i.e. primary vs. secondary) matters for activation and HMWS formation (Daniels et al. 2003). In addition, both the formation of the HMWS and the activation of SgrAI require high concentrations of SgrAI bound to a primary sequence (greater than 1 nM, with 0.9 µM sufficient) possessing sufficient flanking base pairs (Park et al. 2010). Five flanking base pairs found in 18-1 is insufficient for activation (and presumably HMWS) at 37 °C, but 16 bp, such as found in 40-1, PCP, or PC DNA is sufficient for both HMWS at 4 °C and activation of DNA cleavage at 37 °C (Park et al. 2010). The HMWS formed with the shorter 18-1 DNA is likely too unstable at 37 °C but sufficiently stable to form at 4 °C.

To further investigate the HMWS as the activated SgrAI species, we performed a series of DNA cleavage measurements with different reporter DNA constructs containing either a primary site (18-1, 40-1, light and dark blue, Figure 6(B)), a secondary site (18-2, 40-2 A, 40-2B, shades of yellow, Figure 6(B)), or a noncognate site (40-NC, green, Figure 6(B)) with varied concentrations of activator DNA (shown as an increasing grey wedge below the x axis, Figure 6(B)). In this figure, we note that the scale on the y-axis is logarithmic, which allows visualization of rate constants that differ by orders of magnitude. In all cases except 18-2 and 40-NC, an increase in the rate constant was observed with increasing concentrations of activator DNA. This result mirrors the earlier results showing formation of HWMS in native PAGE with increasing concentrations of activator DNA (Figure 5(B)). In the case of 18-2, HMWS were observed in the native PAGE experiments with the addition of PCP DNA (performed at 4 °C, Figure 5(A)), but activation of DNA cleavage (performed at 37 °C) was not found (Figure 6(B)). We reasoned that the shorter flanking base pairs of the 18-2 weakened contacts within the HWMS making it unstable at the higher reaction temperature. Robust activation was however observed with longer flanking DNA, as in 40-2 A and 40-2B (representing the two types of secondary site, 40-2 A: C**C**CCGGTG, 40-2B:**G**ACCGGYG). In the case of the noncognate 40-NC, no activation was observed despite the longer flanking base pairs, consistent with activation and DNA cleavage being specific to primary and secondary sequences (Park et al. 2010; Shah et al. 2015).

Another important observation came from an experiment that involved using an uncleavable DNA construct. In this experiment, the DNA contained a primary sequence embedded in a 22 bp construct, but with a sulfur atom in place of the oxygen atom at the O3' (22-1-3'S). This atom is the leaving group in the DNA cleavage reaction and rendered the DNA uncleavable by SgrAI (Shah et al. 2015). However, this uncleavable primary site containing DNA was found to activate SgrAI to rapidly cleave the externally supplied ^{32}P-18-1 DNA ("+1 µM 22-1-3'S", Figure 6(A)) (Shah et al. 2015). The DNA cleavage rate constant was found to be lower than that seen with 1 µM PCP ($1.47\pm0.12\,min^{-1}$ for the 22-1-3'S vs. $22\pm7\,min^{-1}$ for PCP), although it compared favorably to activation by an unsubstituted version with the same sequence ("22-1", having the rate constant of $1.1\pm0.4\,min^{-1}$, 6th column, Figure 6(A)) (Shah et al. 2015). The lower activation by both 22-1 and its 3'S modified version compared to PCP is likely due to the shorter flanking base pairs rather than the fact it is uncleavable. In conclusion, it was clear that an uncleavable DNA could still serve as an activator of SgrAI, which could lead to *in trans* cleavage of externally supplied DNAs.

In summary, our hypothesis that the HMWS observed in native PAGE is the activated species of SgrAI is supported by the DNA cleavage measurements and show that activation requires a sufficient concentration of SgrAI bound to a primary sequence embedded in a DNA with sufficient flanking base pairs (greater than 5 bp with 7 bp providing some activation and 16 bp providing more). In addition, we showed that this primary sequence need not be cleaved to provide activation. The data also show that DNA with a secondary sequence can be drawn into the HMWS and cleaved at accelerated rates if the enzyme is activated by binding primary DNA sequences containing sufficiently long flanking base pairs.

What is the HMWS?. To further investigate the HMWS, a new series of sedimentation experiments were conducted **(**Figure 7) (Park et al. 2010). In Figure 7(A), the dotted line refers to the c(s) distribution based on the absorbance of the hexachlorofluorescein fluorophore attached to 18-2 (HEX-18-2) DNA in reaction buffer with 10 mM Ca^{2+} and at 4 °C. The species with a sedimentation coefficient of 2S marks the position of the DNA (dotted line, Figure 7(A)). When the DNA was mixed with SgrAI,

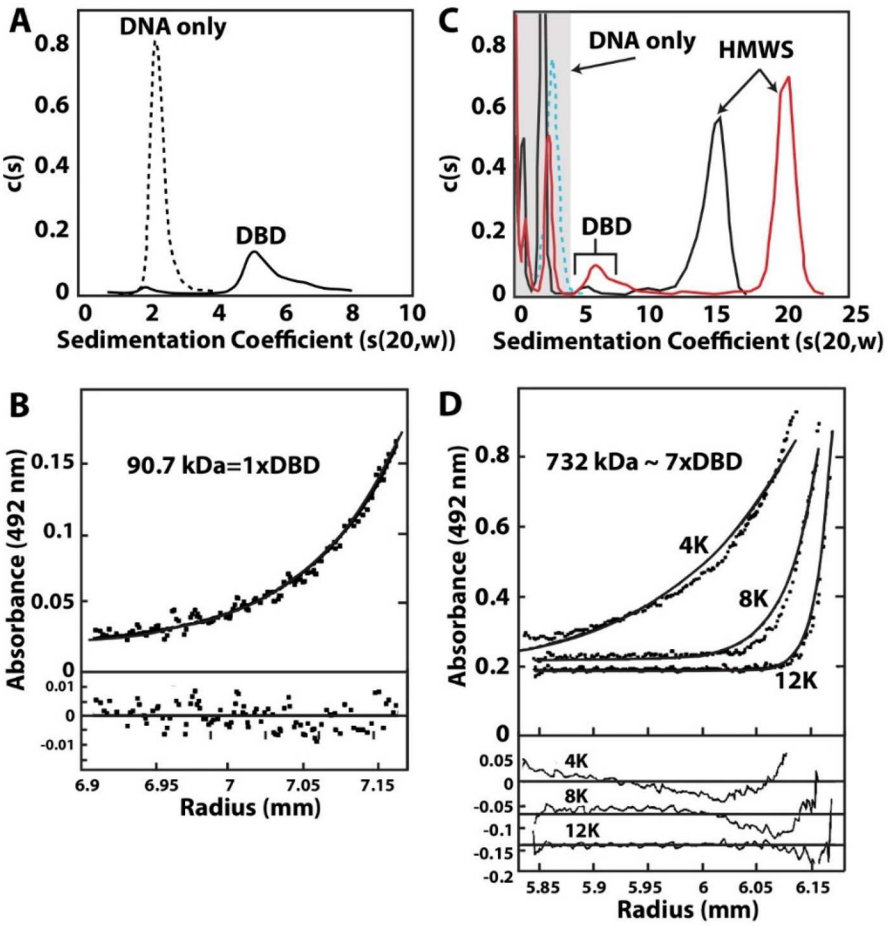

Figure 7. Sedimentation velocity and equilibrium measurements of SgrAI with fluorophore labeled DNA. A. Sedimentation velocity c(s) distribution of a sample of 6 µM SgrAI and 3 µM hexachlorofluorescein labeled 18 bp DNA containing a secondary site (HEX-18-2) at 4 °C. B. Sedimentation equilibrium experiment with 7 µM SgrAI and 3.6 µM HEX-18-2 at 4 °C. The top panel shows data as filled circles and a fit (solid line) to a single species with a molecular mass of 90.7 kDa. The bottom panel shows residuals between the fit and the data. C. Sedimentation velocity experiments with 3 µM fluorescein labeled pre-cleaved primary site embedded in a 40 bp construct (FLO-PCP)(blue dashed line, c(s) values scaled by a factor of 0.5) or with the addition of 6 µM SgrAI in either a low salt buffer (red) or high salt buffer (black) at 37 °C. D. Sedimentation equilibrium experiment with 3 µM SgrAI and 1.5 µM FLO-PCP 37 °C and 4, 8, and 12 krpm rotor speeds. In the top panel, data are shown as filled circles and the line corresponds to a global fit of a single species with a molecular mass of 732 kDa. The bottom panel shows residuals between the fit and data. Residuals in the lower part of panel D have been offset sequentially for clarity by -0.07. (Adapted from Park et al. (Park et al. 2010)).

a larger species was also found with an S value of 5.2 (solid line, Figure 7(A)). To convert S values to molecular weight requires assumptions of molecular shape, therefore the sedimentation equilibrium method (providing shape independent information) was used. Figure 7(B) shows the data for a mixture of 3.6 µM HEX-18-2 and 7 µM SgrAI which data fit well to a model of a monodisperse solution with a molecular weight of 90.7 kDa. This species was therefore assigned as the DNA bound SgrAI dimer (DBD) due its similarity to the expected size of 87.5 kDa.

Sedimentation velocity and equilibrium measurements were also made for SgrAI and the fluorescein labeled pre-cleaved primary site embedded in a 40 bp DNA (FLO-PCP) (Figure 7(C,D)) (Park et al. 2010). The sedimentation experiments were performed at 37 °C and with 10 mM Mg^{2+} to mimic reaction conditions, and absorbance of the fluorophore used to detect species in solution. The sedimentation velocity results recapitulated the presence the unbound DNA and DBD, as was observed with HEX-18-2, but an additional larger species was also found with an S value of ~15-20 (HMWS, Figure 7(C)). The S value of this species varied with buffer conditions. For example, at higher ionic strength (150 mM NaCl, black line, Figure 7(C)), the HMWS presented with a S value of ~15, but with lower ionic strength buffer (50 mM potassium acetate, red line, Figure 7(C)), the HMWS was larger (~20 S, red line, Figure 7(C)). To gain some idea of the exact size of the HMWS, sedimentation equilibrium measurements were performed under the lower ionic strength conditions (Figure 7(D)). The data (points, Figure 7(D)) did not fit

well to a single species (black lines, upper panel, Figure 7(D)) and the residuals from the fit showed systematic deviations (lower panels, Figure 7(D)) indicative of the presence of multiple species bearing different molecular weights. A fit to the single species gave a value of 732 kDa, suggestive of an average molecular weight of species in solution corresponding to approximately 7 copies of the DBD (Park et al. 2010). From the sedimentation data, it was apparent that the HMWS was polydisperse, and likely composed of multiple copies of the DNA bound SgrAI.

The SgrAI/PC DNA sample was next subjected to both conventional native mass spectrometry as well as ion mobility mass spectrometry (IMMS) which is a native MS approach that separates species in flight during the electrospray process based on their size (Ma et al. 2013; Christofi and Barran 2023). Samples thus analyzed contained high concentrations of SgrAI and DNA (10-12 µM SgrAI and 1:1 to 1:4 SgrAI:DNA ratios) in a buffer containing 100 mM ammonium acetate, 5 mM calcium tartrate, and 1 mM DTT, and at ambient temperature (with the ion zone maintained at 30 °C). First, the spectrum was measured for SgrAI without DNA (Figure 8(A)). As expected, a species with the molecular mass of the SgrAI dimer was found to be the dominant form. Next, the spectrum of SgrAI bound to a 40 bp DNA containing a secondary sequence (40-2) was measured (Figure 8(B)), and in agreement with the native PAGE and sedimentation studies, most of the sample was found to be in the form of a single DNA bound SgrAI dimer (DBD), although a minor species containing 2 copies of DBD (2xDBD, Figure 8(B)) was also identified. Next, the spectrum of SgrAI with PC DNA (a pre-cleaved 40 bp DNA containing a single primary site) was measured (Figure 8(C)). In this case, species larger than 1xDBD dominate the spectrum, including species identified as containing 2xDBD, 3xDBD, as well as a large unresolved set of peaks at higher m/z (labeled as HMWS, Figure 8(C)). Ion mobility mass spectrometry was then used to attempt to resolve overlapping peaks and identified six clusters or zones (I-VI, Figure 8(D)). Peaks within Zones I-III were found to originate from species containing 1x, 2x, or 3xDBD, respectively. Follow-up measurements identified several larger species in Zones IV-VI, including a 6xDBD complex in Zone VI (Figure 8(E)), 10xDBD in Zone IV, and 19xDBD in Zone V.

There were several conclusions drawn from this set of experiments. First, it became apparent that the

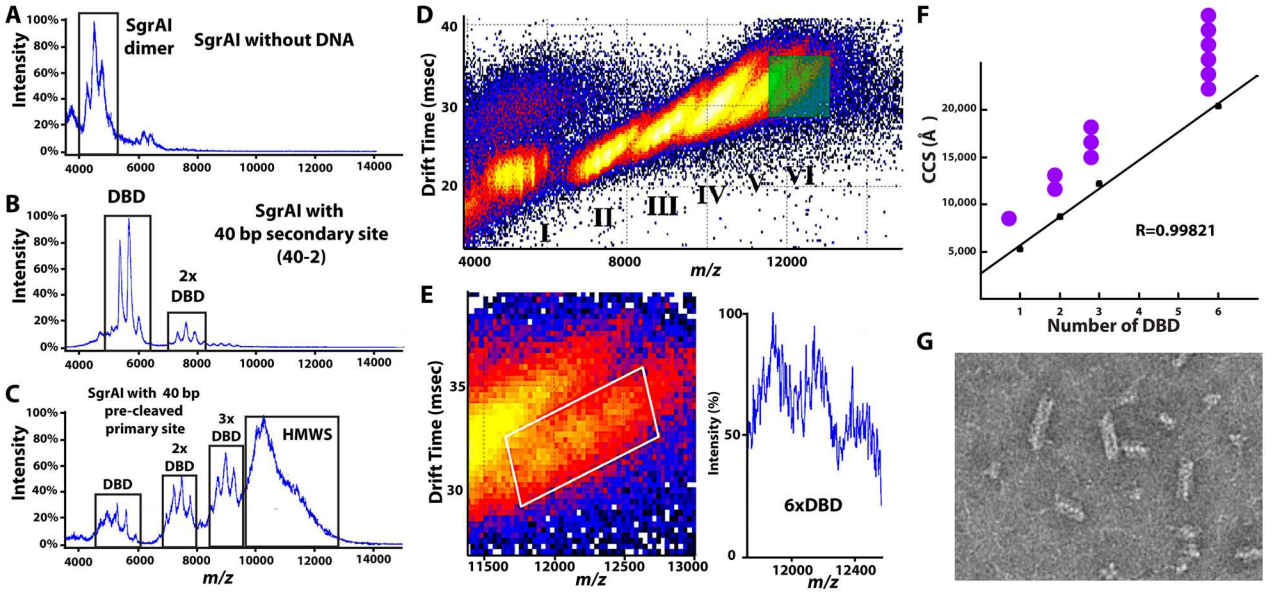

Figure 8. Characterization of DNA bound SgrAI using mass spectrometry (MS) and TEM. A. MS spectrum of SgrAI without DNA. Peaks indicate a single SgrAI dimer. B. MS spectrum of SgrAI with a 40 bp DNA containing a secondary sequence (40-2). Peaks correspond to mostly a single DBD (DNA bound SgrAI dimer) with the presence of a minor amount of two DBD bound in a complex (2xDBD). C. MS spectrum of SgrAI with a twofold molar ratio of PC DNA (pre-cleaved primary site embedded in a 40 bp DNA). Peaks were identified corresponding to a single DBD, two DBD bound in a complex (2xDBD), and three DBD bound in a complex (3xDBD). Overlapping peaks were found in the region of 10-13 km/z and were subsequently identified as originating from the HMWS. D. Ion mobility mass spectrum of SgrAI with PC DNA showing the separation of species into six zones, I-VI. E. Close-up of Zone VI (left) in the IMMS data and corresponding mass spectrum (right) showing peaks corresponding to 6xDBD. F. Plot of the collisional cross-section (CCS) derived from IMMS data vs. corresponding number of DBD present showing a linear relationship. G. Negative stain TEM of SgrAI/PC DNA showing the presence of species with a regular, repeating structure. (panels A-E adapted from ma et al. (Ma et al. 2013), panel F adapted from Shah et al. (Shah et al. 2015), panel G from Lyumkis et al. (Lyumkis et al. 2013)).

species dominating the spectra contained repeating integer numbers of SgrAI DBDs, indicating that the DBD constitutes the baseline building block of the HMWS. Second, there was also an apparent periodicity of ~4 DBDs. This was suggested based on the observation that the m/z of species with increasing number of DBD increases regularly until about 4xDBD, whereas species with numbers of DBD from 4-19 exhibit similar m/z. The charge z of a species is proportional to its exposed surface area; species with DBD from 1-4 exhibit increasing m/z, meaning decreasing surface area per DBD in the complex. After ~4 DBD, additional DBD copies change the exposed surface area by a regular value, which is why all species larger than ~4 DBD exhibit roughly the same m/z. Another measurement that can be extracted from the IMMS data that suggests a regular, repeating structure of DBD in the HMWS is the collisional cross-section (CCS) of the molecular complexes. Figure 8(F) shows that a linear relationship between the CCS and the number of DBD present in the species exists. The simplest type of assembly which is both large and heterogeneous in mass but contains a regular repeating structure is a filament (symbolized by purple filled circles, Figure 8(F)). Direct confirmation of this hypothesis was then shown using electron microscopy of a negative stained sample of SgrAI and PC DNA (Figure 8(G)). The image shows the clear presence of species of different sizes but containing the same regular, repeating filamentous structure as suggested by the IMMS data.

The HMWS of SgrAI bound to DNA is a helical filament

To determine the molecular architecture and the mechanism of multimerization of the HMWS, we turned to single-particle electron microscopy (EM). Single-particle EM techniques are excellent for identifying and resolving structurally heterogeneous assemblies, including when different specimen compositions are present within a sample, as was expected based on the results of native MS. Class averages and corresponding 3D reconstructions from EM images of negatively stained HMWS assemblies revealed multiple distinct species containing anywhere from 1 to 13 DBD (Figure 9). The presence and/or absence of a single DBD clearly distinguished each successive reconstruction (Figure 9), just as the earlier IMMS data predicted (i.e. that the DBD was the building block of the filament). These early reconstructions were clearly defined by left-handed helical organization, and it was possible to observe the helical periodicity of ~4 DBD per helical turn, also predicted by the IMMS data. The larger filamentous assemblies were next subjected to cryo-electron microscopy (cryo-EM) to obtain higher-resolution reconstructions. The cryo-EM structure was solved initially to 8.6 Å resolution, and subsequently to 2.5 Å, using helical refinement techniques and confirmed the left-handed helical nature of the filament with an 85.8° rotation and 21.2 Å translation between successive DBD (Figure 10(A)) (Lyumkis et al. 2013; Polley et al., 2019; Shan et al. 2024). Each DNA bound SgrAI dimer (DBD) is positioned with its DNA binding cleft and N-terminal protrusions (those that were additions to SgrAI compared to NgoMIV and Bse634I, red ovals, Figure 3) facing inward making a tight protein-protein interface (Figure 10(A,B)). The corresponding face of the SgrAI dimer that is used at the dimer-dimer interface in the tetramers of NgoMIV and Bse634I is facing away from the helical axis (yellow circle, Figure 10(A)). The base pairs flanking the recognition site on either side make contacts to neighboring DBD (green box, Figure 10(C), see

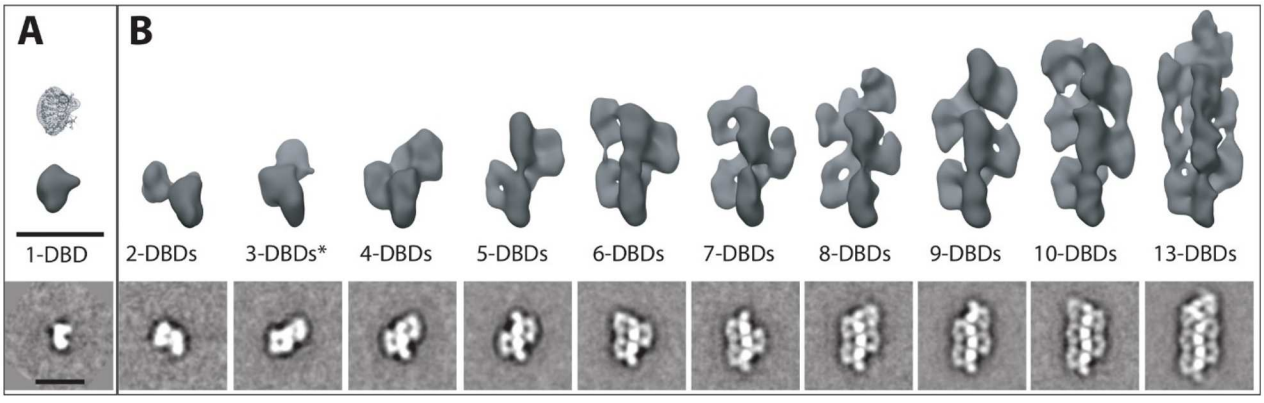

Figure 9. 2D Class averages and 3D reconstructions of the SgrAI/DNA filament from negative stain TEM. A. One DNA bound SgrAI dimer (DBD). (lower) 2D class average for an isolated DBD. The scale bar represents 150 Å. (Middle) 3D reconstruction of one DBD created by the random conical tilt method. The scale bar represents 150 Å. (Upper) X-ray crystal structure of DNA bound SgrAI fit into the 3D reconstructed map. B. As in A, particles with 2-13 DBD. (from Lyumkis et al. (Lyumkis et al. 2013)).

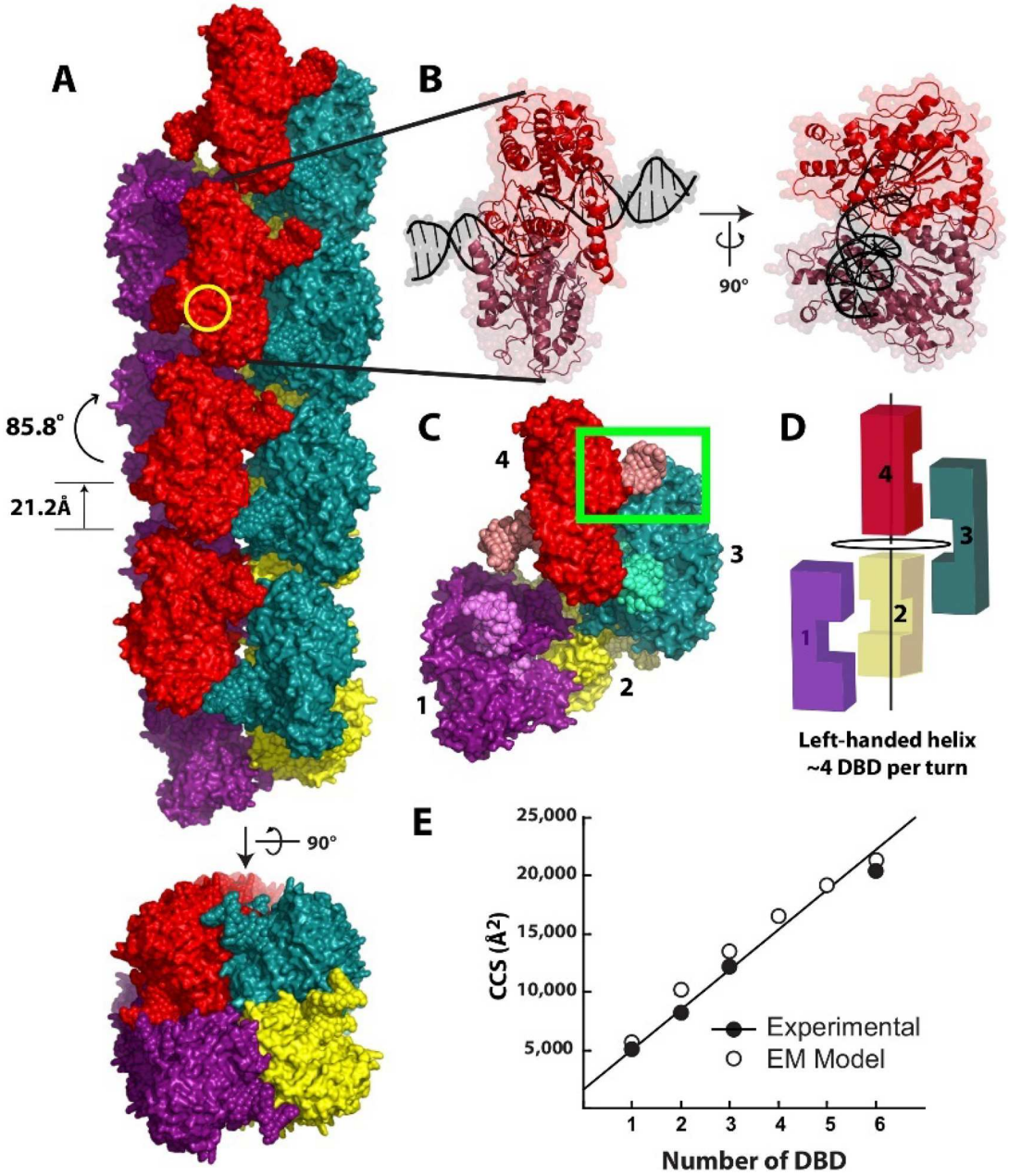

Figure 10. The SgrAI/DNA filament structure and correlation with IMMS data. A. Model of the SgrAI/DNA filament showing a left-handed helix with 85.8° rotation and 21.2 Å translation per DBD. Yellow circle identifies the location of an E301W mutation. The green box identifies contacts between the flanking DNA of one DBD and the SgrAI of a neighboring DBD. B. Ribbon diagram of a DBD from the filament. C. Four DBD of the filament showing its left-handed helical nature. D. Schematic of DBD arrangement shown in in panel C. E. Correlation of SgrAI/DNA filament model shown in a with collisional cross-sections measured by ion mobility mass spectrometry. (panels A-D from Shan et al. (Shan et al. 2022). Panel E adapted from Lyumkis et al. (Lyumkis et al. 2013)).

also Figure 11(A) for close-up), which explained earlier observations on the importance of DNA length and flanking bp on SgrAI activation and HMWS stability. The SgrAI/DNA filament also exhibits the same calculated collisional cross-section (CCS) that was measured previously using ion mobility mass spectrometry (Figure 10(E)).

To investigate whether the structure determined by cryo-EM is indeed the activated species of SgrAI, a series of mutations at the protein-protein and protein-DNA interfaces between adjacent DBD in the filament structure (Figure 11(A)), as well as one distant from the interface but at the corresponding tetrameric interface of NgoMIV and Bse634I (yellow circle, Figure 10(C)), were created and tested for their effects on SgrAI activation (Figure 11(B)) (Park et al. 2010; Shah et al. 2015; Barahona et al. 2019). Most mutations did not disrupt DNA binding or the basal, unactivated DNA

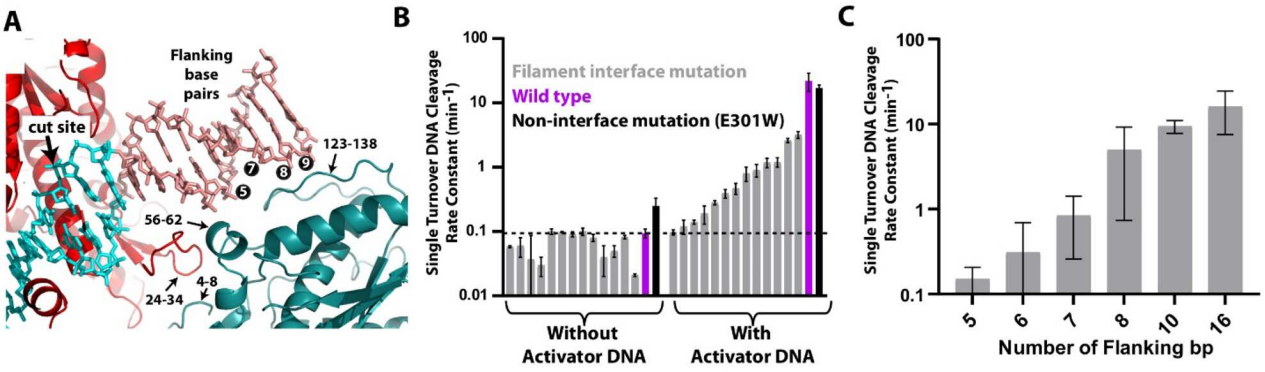

Figure 11. Investigation of the protein-protein and protein-DNA interfaces between adjacent DBD in the cryo-EM filament structure. A. Protein-protein and protein-DNA interfaces between adjacent DBD in the filament structure determined by cryo-EM. The position of residues identified for mutation (black) as well as the position of flanking base pairs 5-9 (white in black circles). This is a zoom in on the green boxed region of Figure 10(C). B. The single turnover DNA cleavage rate constants of wild type (purple bars) and mutant (grey and black bars) SgrAI on ^{32}P-18-1 in the absence of added activator DNA (left) or presence of 1 µM PC DNA (right). Note the log scale on the y axis. The dotted line indicates basal, unactivated activity. Rate constants for SgrI with mutations at residues at the protein-protein or protein-DNA interfaces in the filament structure are shown in grey, and that for a single mutation (E301W) at the putative NgoMIV-like tetramer interface is shown in black. All mutations at the inter-DBD interfaces found in the filament affect the activation of SgrAI without affecting its basal DNA cleavage rate. E301W had no effect on either basal or activated rate constants for DNA cleavage consistent with this region being far from dimer-dimer interactions as seen in the filament structure. C. Tests for SgrAI activation by DNA constructs containing a primary sequence and varied numbers of flanking base pairs. Note the log scale on the y axis.

cleavage rate of SgrAI, yielding similar DNA cleavage rate constants similar to wild type SgrAI in the absence of activator DNA ("Without Activator DNA", wild type SgrAI shown in purple, the dotted line represents the unactivated DNA cleavage rate constant, Figure 11(B)). Next, the degree to which the mutations disrupted filament formation was investigated by measuring the activated rate constant for DNA cleavage ("With Activator DNA", Figure 11(B)). All mutations found at the protein-protein and protein-DNA interfaces between adjacent DBD in the filament structure showed lessened activation, though to varying degrees. In contrast, a mutation far from the interface (E301W), but at the canonical type IIF RE tetrameric interface (black, Figure 11(B)) had no effect on the activation of SgrAI.

To investigate the protein-DNA interactions between each DBD and the flanking DNA of a neighboring DBD in the filament, primary sites with different flanking DNA lengths were also tested (Figure 11(C)) (Lyumkis et al. 2013). The DNA constructs contained a primary sequence and 5-8, 10, or 16 bp flanking the recognition sequence on either side. Note the log scale on the y axis showing the activated rate constant for cleavage of 1 nM ^{32}P-18-1 with 1 µM SgrAI and 1 µM of each of the different DNAs. The cleavage rate constant is very slow for those constructs with flanking bp of 5-7 but increases with increasing flanking bp length until a maximum between 8 and 10 bp (Figure 11(C)). These DNA cleavage activity measurements agree with the filament structure; Figure 11(A) shows that the flanking bp forming important contacts to neighboring DBD in the filament occur at base pairs 5-9 bp from the recognition sequence (white text in black circles, Figure 11(A)).

Mechanism of SgrAI activation by filamentation

Comparison of the SgrAI/DNA structure in the filamented (Polley et al., 2019) and unfilamented (Dunten et al. 2008) state shows an ~11° rotation of one subunit of the SgrAI dimer relative to the other in a direction nearly perpendicular to the helical axis of the bound DNA (Figure 12(A)). To accommodate this conformational change, residues at the dimeric interface shift, which propagates to the protein-DNA interface that includes the active site (Figure 12(B), red arrow, Figure 12(C)). As a direct consequence, the binding of the site B metal ion is now stabilized *via* hydrogen bonding of a metal ion coordinated water to the carbonyl oxygen of Thr186 (light blue sphere, Figure 12(C)). As described above, divalent cation dependent endonucleases like SgrAI are considered to function *via* a "Two-Metal Ion Mechanism", where two divalent metal cations (typically Mg^{2+}) are positioned around the scissile phosphate, as shown in Figure 3(C). The structure of SgrAI bound to DNA (primary or secondary) in its non-filamented dimeric form showed occupation of only site A (as well as at a distal site D) but not site B (Figure 3(D)). As predicted earlier, a

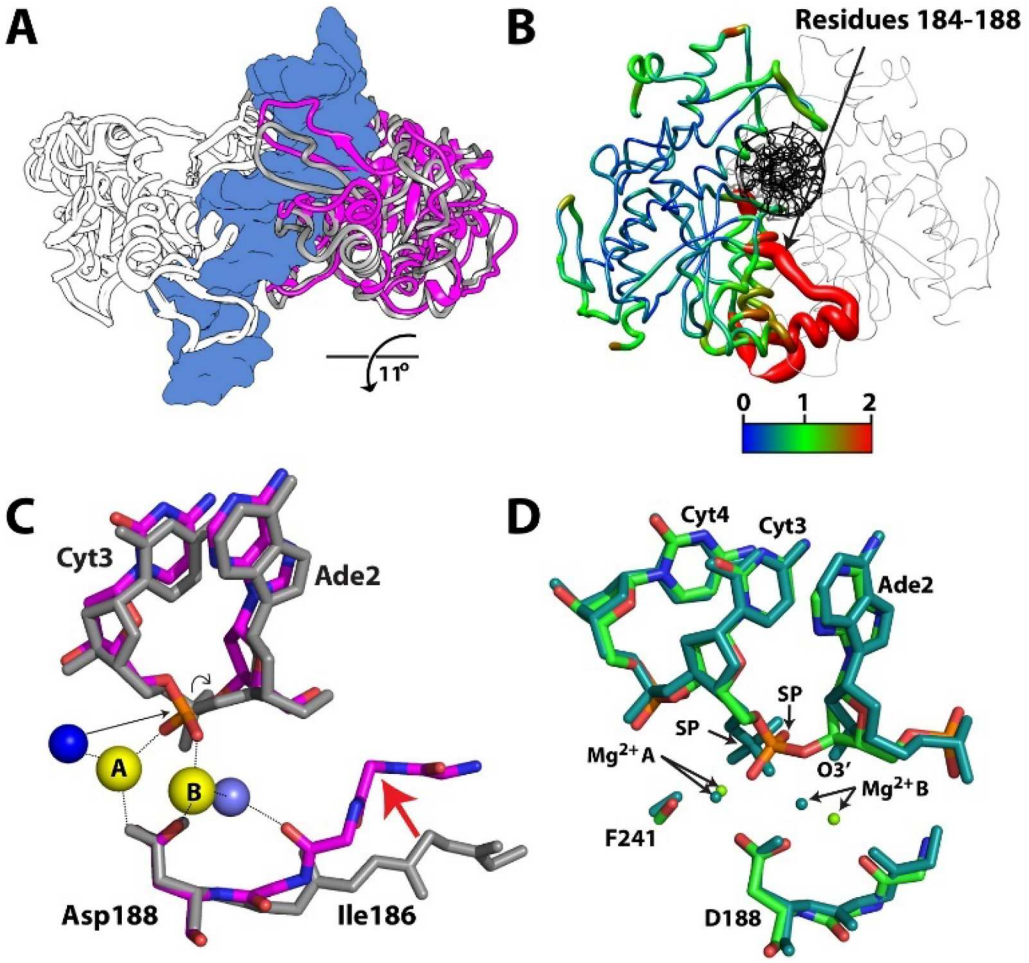

Figure 12. Conformational changes in the DBD in the filamentous form propagate to the active site to create a metal ion binding site B. A. Superposition using the left subunit of the low activity L state of SgrAI (white and grey, PDB code 3DVO) and the high activity H state from filamentous SgrAI (white and magenta, PDB code 7SS5) to show the change in position of the subunit on the right. The subunit rotates 11° about an axis nearly perpendicular to the axis of the bound DNA (blue). (Adapted from Ghadirian et al. (Ghadirian et al. 2024)), B. Cartoon of SgrAI dimer emphasizing the shifts in amino acid positions within one subunit (RMSD by color, in Å) in comparing the L and H states. (Adapted from Shan et al. (Shan et al. 2022)), C. Comparison of L and H states using a superposition of all atoms of the subunit shown identifying the change in the segment between residues 184 and 187 (red arrow). Two Ca^{2+} bind in sites A and B. The direction of nucleophile attack is shown by a straight arrow and bond breakage is shown by a curved arrow. D. Superposition of two H state (filamentous) structures of SgrAI bound to Mg^{2+}. The wild type structure (PDB code 3GBI) is shown in dark green, and the scissile phosphate is found cleaved. The structure of and active site mutant bound to uncleaved DNA is shown in light green (PDB code 3GBJ). (Adapted from Shan et al. (Shan et al. 2024)).

conformational change occurs in SgrAI to stabilize the metal ion binding at site B. This conformational change is stabilized by contacts to neighboring DBD in the filament (Figure 10(A)). Modeling shows that only the high activity (H) conformation of SgrAI bound to DNA is compatible with the filament structure as the low activity conformation (L state, present in the non-filamentous form) disrupts contacts between DBD and introduces destabilizing steric conflicts (Polley et al., 2019).

Taken together, the results of all biochemical and structural studies were combined to produce the mechanistic model for DNA cleavage by SgrAI shown in Figure 13. In panel 1, the low activity (L) state that is observed in the crystal structures of nonfilamented SgrAI show occupation of only the site A Mg^{2+} (green sphere). In panel 2, the activated or high (H) activity conformation which is observed in cryo-EM reconstructions of filamentous SgrAI bound to DNA exhibits a shift of the segment containing Thr186 closer to the DNA. This shift stabilizes the binding of Mg^{2+} in site B *via* hydrogen bonding of the backbone carbonyl of shifted Thr186 with a water Mg^{2+}-coordinated water molecule (light blue, Figure 13). In this state, another water molecule coordinated by the Mg^{2+} ion in site A (shown as hydroxide, dark blue, Figure 13) is positioned to attack the phosphorus atom of the scissile phosphate (see red arrow in panel 2, Figure 13),

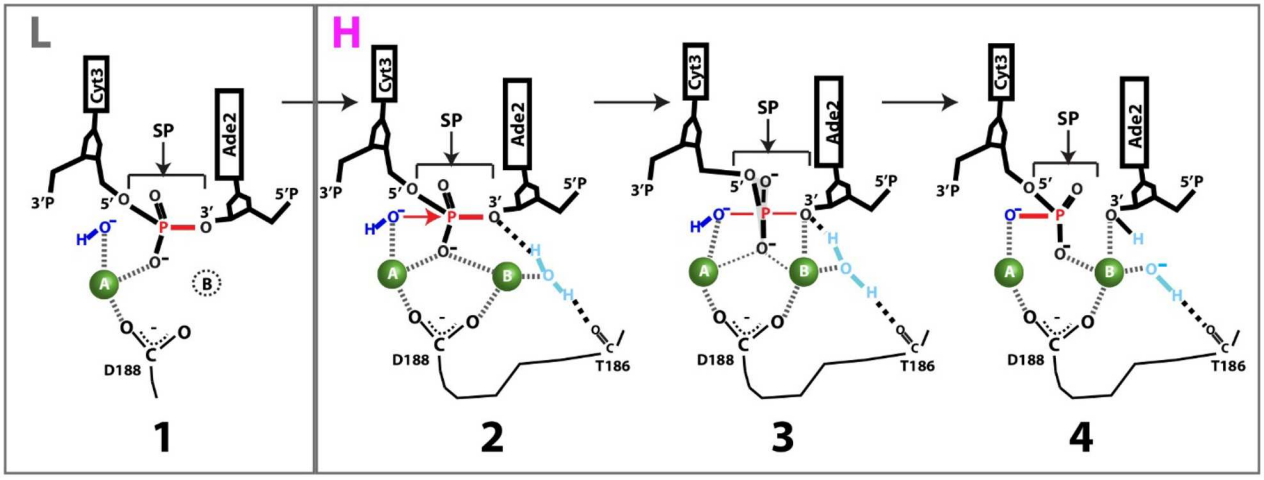

Figure 13. Active site mechanism of DNA cleavage by activated SgrAI. Current model for the activated DNA cleavage mechanism of SgrAI based on all structural studies. State 1: the low activity, non-filamentous, or L state. A single Mg^{2+} is bound in site A (green sphere). State 2: Just prior to nucleophilic attack (red arrow) of the scissile phosphate (SP) phosphorus atom by an activated water (shown as hydroxide, dark blue) coordinated to the site A Mg^{2+}. SgrAI is in the high activity (H) state found in filamentous SgrAI. A segment containing Thr186 is shifted closer to the SP and stabilizes binding of Mg^{2+} to site B *via* hydrogen bonding to a coordinated water molecule (light blue). State 3: Model of the transition state (or short-lived intermediate) after nucleophilic attack and prior to bond breakage. The pentacoordinate phosphorus atom exhibits trigonal dihedral symmetry with the leaving group (O3′) directly opposite of the attacking group. State 4: after bond cleavage. (Adapted from Shan et al. (Shan et al. 2024)).

thereby producing the pentacoordinate species seen in panel 3. This pentacoordinate species may be a high energy intermediate or a true transition state (Gerlt 1993; Cassano et al. 2004; Kamerlin and Wilkie 2007; Lassila et al. 2011). Its collapse into the more canonical tetrahedrally coordinated phosphate, shown in panel 4, involves the breaking of the P-O3′ bond (red in panels 1-3) to produce the active site arrangement seen in the structure of filamentous SgrAI bound to Mg^{2+} and cleaved DNA (Shan et al. 2024).

Control of the DNA cleavage reaction in our model occurs *via* controlling the position of the divalent cations. The exact roles of divalent cations in DNA cleavage reactions have been the subject of many investigations and discussions (Pauling 1947; Burgers and Eckstein 1979; Brody and Frey 1981; Aggarwal 1995; Horton et al. 1998; Horton and Perona, 1998, 1998; Viadiu and Aggarwal 1998; Martin et al. 1999; Mizuuchi et al. 1999; Horton et al. 2000; Horton and Cheng 2000; Sam et al. 2001; Horton et al. 2002, 2002a, 2002b; Chevalier et al. 2004; Etzkorn and Horton, 2004, 2004; Horton and Perona 2004; Lee et al. 2005; Nowotny et al. 2005; Joshi et al. 2006; Nowotny and Yang 2006; Yang et al. 2006Steitz and Steitz, 1993; Babic et al. 2008; Dunten et al. 2008; Horton 2008; Little et al. 2008; Yang 2011; Polley et al., 2019). In general, they are thought to perform numerous critical functions for the enzyme, such as activation of the nucleophile, stabilization of the transition state, and stabilization of the leaving group. They do this *via* direct coordination to the oxygen atoms from the water molecules and the DNA backbone. In Figure 13, the site A Mg^{2+} is seen to coordinate the nucleophile (water or hydroxide) and a non-esterified oxygen (pro-S_P) of the scissile phosphate. These interactions can be important for numerous processes, including: 1) activation of the nucleophile *via* stabilization of the deprotonated water molecule, i.e. hydroxide, 2) optimally positioning both the nucleophile and the scissile phosphate for the enzymatic reaction (the nucleophile, phosphorus, and O3′ leaving group should be roughly in-line, and the distance between the nucleophile and the phosphorus atom should be near the van der Waals distance, i.e. ~3.3 Å), 3) screening of the negative charge on the scissile phosphate to facilitate the close approach by the nucleophile, and/or 4) stabilization of the transition state, which is expected to possess two negative charges following nucleophilic attack. Important as these functions are, occupation of site B is also necessary in our model to achieve full DNA cleavage rates. The postulated functions of the site B Mg^{2+} partially overlap with those of the site A Mg^{2+} and include: 1) positioning of the scissile phosphate, 2) stabilization of the transition state, and 3) stabilization of the O3′ leaving group. The latter function is necessary because a negative charge on the O3′ would be very unstable; stabilization could occur either *via* protonation by a Mg^{2+} coordinated water molecule or *via* direct ligation to the site B Mg^{2+} (with subsequent protonation). Although we do not see direct coordination

of a Mg^{2+} coordinated water molecule to the O3′ in any structure solved to date, superposition of two structures containing uncleaved DNA, one with Ca^{2+} (PDB code 7SS5) and one with an active site mutation and Mg^{2+} (PDB code 3GBJ) shows a water molecule coordinated to the site B Ca^{2+} that is close enough to make a hydrogen bond to the O3′ within the mutant structure. Unfortunately, the resolution of the mutant structure is insufficient to identify the positions of water molecules. However, since the position of the O3′ in the mutant structure (with bound Mg^{2+}) exhibits an orientation of the scissile phosphate and O3′ more similar those found in other restriction endonucleases, we include this aspect, namely protonation of the O3′ with a site B coordinated water, in our model (panel 4, Figure 13). Finally, since we see direct coordination of the O3′ following bond cleavage in our structure with cleaved DNA and Mg^{2+} (PDB code 3GBI, dark green, Figure 12(D)), we also include direct ligation of the site B Mg^{2+} to the O3′ as a means to stabilize the leaving group in our model of the reaction (panels 2-3, Figure 13).

Kinetic model of activated DNA cleavage by filamentous SgrAI

The development of a computational mechanistic model

The studies described above indicate that SgrAI forms polymeric filaments as part of its DNA cleavage mechanism. Because this is a new paradigm in enzyme regulation, there are fundamental questions for how filamentation regulates the catalytic properties of the enzyme and the substrate. We summarize these in Table 1. To address these questions, and to develop a coherent understanding of enzyme regulation *via* filamentation, computational models of the DNA cleavage pathway were built and tested against DNA cleavage reaction data.

To develop and refine the model, it was necessary to develop enzymatic assays that measure diverse steps in the reaction and in which the substrates can be modulated to yield readouts of the SgrAI-mediated DNA cleavage activity. Three different assays were therefore developed and used for the DNA cleavage measurements, outlined in Figure 14. In the first approach, a single turnover enzymatic activity assay was used to measure the DNA cleavage rate using a low (1-150 nM) concentration of ^{32}P-18-1 as a reporter for DNA cleavage, unlabeled activator PC DNA at different concentrations to control the level of filamentation, and high concentrations of SgrAI (1-2 µM) to saturate binding of all DNA molecules in the reaction (Figure 14(A)). This reaction, which was previously introduced in the context of Figure 6, measures the enzymatic steps from DNA binding to DNA cleavage but is not influenced by filament dissociation or the release of the cleaved DNA (i.e. product release). In addition, two other experimental approaches which made use of the FRET pairs fluorescein (Flo) and rhodamine-X (Rox) attached to the DNA substrates were used. In one version of the assay, a doubly labeled reporter DNA (Flo-18-1-Rox, Figure 14(B)) was used. When duplexed, the emission from Flo is quenched by the close proximity of Rox. However, after DNA cleavage and release, the experimental conditions are such that the cleaved duplexes dissociate into single strands, leading to a loss in FRET and unquenching of the Flo emission. This approach also measures all steps from DNA binding but includes the two additional steps of filament dissociation and product release (Figure 14(B)). Finally, in Figure 14(C), the assay with Flo on the activator DNA and Rox on the reporter DNA is described. In this case, the Flo signal is quenched upon filament association and unquenched upon filament dissociation. Hence these three types of assays probe different combinations of reaction steps, which aide in constraining rate constants during global

Table 1. Questions to be addressed by the kinetic study of DNA cleavage by SgrAI.

1) What controls filament length? Can the filament grow infinitely long?
2) Is the cleaved DNA released directly from the filament, or must SgrAI dissociate from the filament to release its bound cleaved DNA?
3) Does the cleaved DNA become trapped in filaments?
4) If DBDs dissociate from filaments to rapidly release cleaved product DNA, how can it reside long enough in filaments for DNA cleavage to occur?
5) Does the affinity of DBDs for filaments depend on filament length?
6) Do filaments assemble through the bidirectional addition of single DBD to either end only, or can filaments of varying lengths come together to generate longer filaments?
7) Conversely, when filaments disassemble, do DBDs dissociate individually from the ends, or can filaments break apart anywhere along their length?
8) Does SgrAI bound to cleaved DNA associate with filaments with the same affinity and kinetics as SgrAI bound to uncleaved DNA? Alternatively, is DNA cleavage a trigger for dissociation of SgrAI from the filament?
9) Do all DBDs within a filament cleave their bound DNA in a coordinated manner, or is cleavage of DNA independent in different DBDs within the same filament?
10) Why is the observed DNA cleavage rate constant from single turnover reactions dependent on the concentration of activator DNA? Could this dependency be due to the rate-limiting association of DBD into the filament?
11) Why is it beneficial for bacterial species to evolve a mechanism that involves enzyme filamentation? How does the filamentation mechanism translate into rapid cleavage of invading phage DNA?

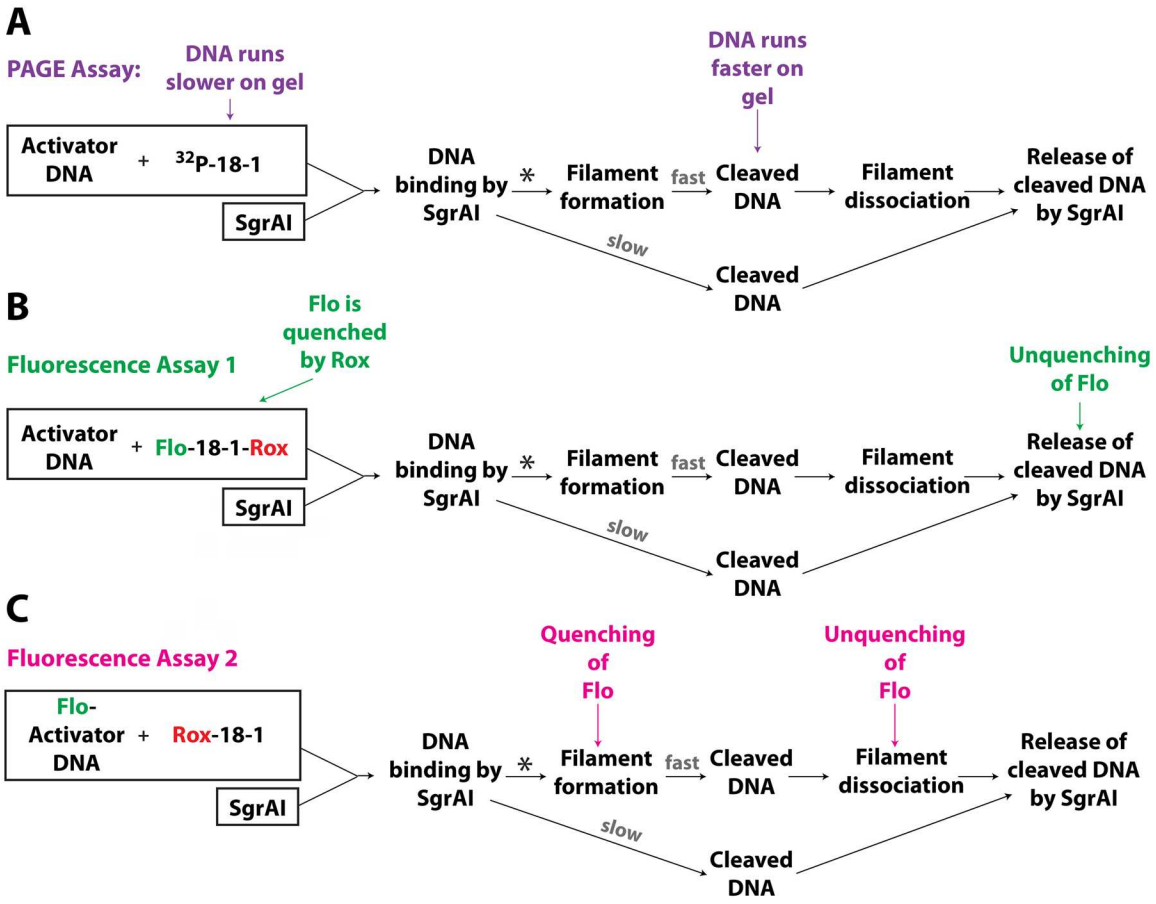

Figure 14. Experimental setup for the three distinct DNA cleavage assays that provided the data for global kinetic modeling. A. 1-150 nM ^{32}P labeled DNA 18 bp DNA containing a primary sequence (^{32}P-18-1) and varied concentrations of activator DNA (PC DNA, from 0-1 μM) were mixed with 0.3-2 μM SgrAI in buffer containing 10 mM Mg^{2+} at 25 °C to initiate the DNA cleavage reaction. Denaturing PAGE followed by autoradiography of cleaved and uncleaved ^{32}P-18-1 provided measurements of all reaction steps up to and including DNA cleavage. B. As in A, however the reporter DNA is labeled with both fluorescein (Flo) and rhodamine-X (Rox) (Flo-18-1-Rox). Detection of the unquenching of the fluorescein emission allowed measurement of steps from DNA binding through to the release of cleaved DNA.C. A third scenario where Flo is placed on the activator DNA and Rox is placed on the reporter. The assembly of filaments is detected by the quenching of the Flo signal, while filament dissociation is detected by its unquenching. *indicates the rate limiting step of filament association which is modulated by the concentration of activator DNA. (Adapted from Ghadirian et al. (Ghadirian et al. 2024)).

data fitting of the different reaction steps. Multiple reaction measurements were made using each approach and differing concentrations of activator DNA which strongly influenced the overall rate by affecting the rate of filamentation (indicated by a "*" in Figure 14).

To test and validate the approach used in Figure 14(C), and further investigate filament formation by SgrAI, a titration experiment was performed using 50 nM of Rox-18-1, 2 μM SgrAI, and varied concentrations of Flo-PC (Figure 15(A)) (Park et al., 2018). Because Flo-PC is expected to bind to SgrAI in a 2:1 complex, the concentrations in Figure 15(A) are given in terms of the SgrAI/Flo-PC_2 complex (a.k.a EP_{40}). Ca^{2+} was used in place of Mg^{2+} to stall the DNA cleavage reaction, such that binding to the uncleaved Rox-18-1 could be measured. The emission of Rox (with excitation of Flo) increases with increasing concentrations of PC DNA until it reaches a maximum at approximately 800 nM Flo-PC (400 nM Flo-EP_{40}), presumably due to the association of both Flo and Rox DNA in the filament, thereby allowing for FRET to occur. Figure 15(B) plots the maximum emission as a function of [Flo-EP_{40}] along with a fit to the Hill equation. The $K_{1/2}$ of binding was found to be 0.16 ± 0.03 μM and the Hill coefficient, a measure of cooperativity in binding, was found to be 1.1 ± 0.1. This $K_{1/2}$ is consistent with the native PAGE titrations of HMWS (Figure 5(B,C)), and the Hill coefficient of ~1 indicates that DBD do not assemble cooperatively into the filament, consistent with the observation that each DBD contacts only the DBD before and the DBD after it in the filament (Figure 10(A)). This also predicts that DBD should bind filaments with the same affinity,

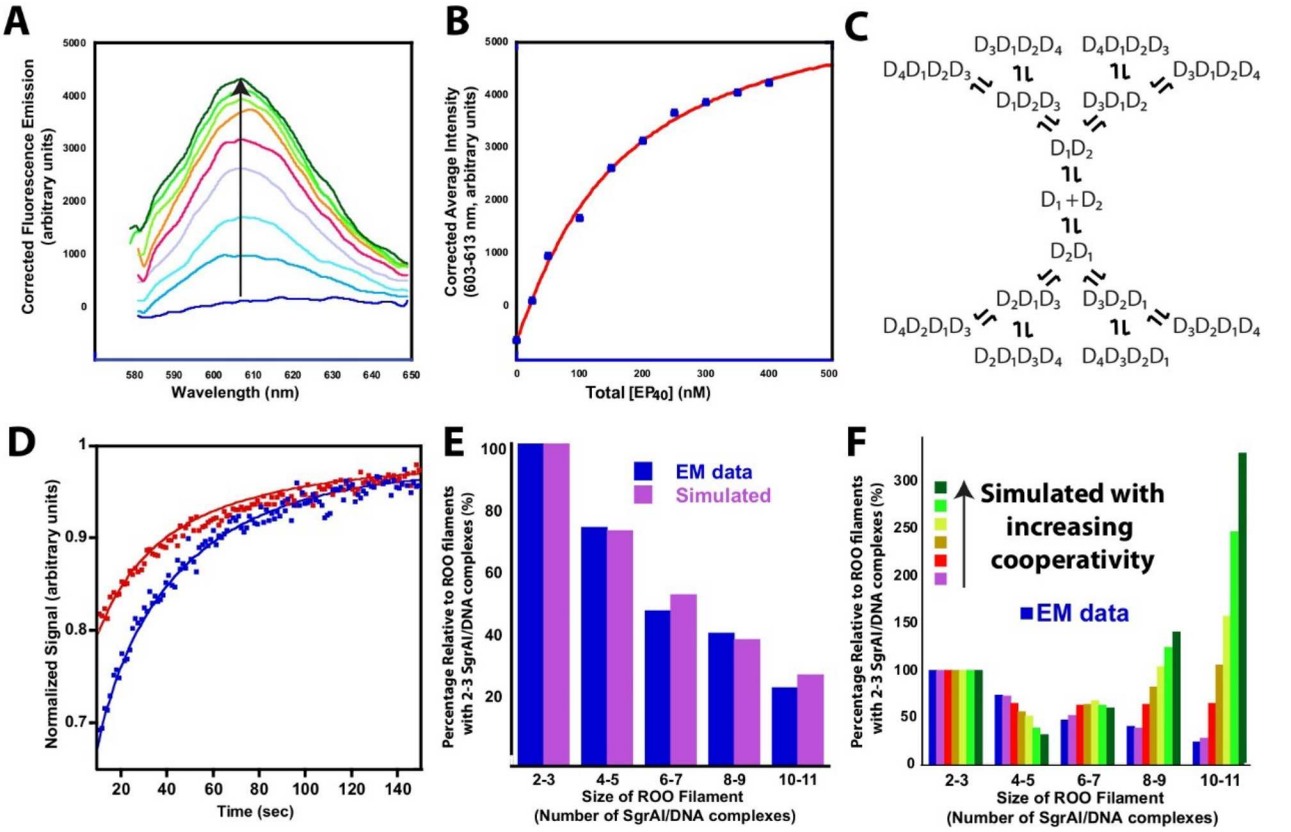

Figure 15. Experimental and simulated assembly of filaments from SgrAI bound to uncleaved (ES_{40}) or cleaved (EP_{40}) 40-1 DNA. A. FRET titration of Rox-PC with Flo-PC in the presence of excess SgrAI. The arrow shows how the FRET signal increases with increasing concentration of activator PC DNA. B. Data from A (blue circles) fit to the Hill equation (blue line) giving a $K_{1/2}$ of 0.16 ± 0.03 µM, and Hill coefficient of 1.1 ± 0.1. C. Schematic of model used in the global data fitting. Each step has a corresponding forward and reverse rate constant to be fit by global data fitting. D. Example data (blue and red points) and simulations from the calibrated computational model (blue and red lines). E. Simulation using filaments up to 12 DBD in size and rate constants derived from global data fitting. The distribution of filament sizes from the simulation (purple) is shown compared to that determined experimentally from negative stain TEM images (blue). F. As in E, however with increasing degrees of cooperativity added showing the decrease in agreement with the experimental data when increasing degrees of cooperativity are used in the simulations. Cooperativity was modeled as a decreasing dissociation rate constant for DBD from filaments of increasing size. (Adapted from Park et al. (Park et al., 2018)).

regardless of filament length, which addresses Question #5 in Table 1.

In order to extract the rate constants for all steps of the reaction pathway from the data collected using the three approaches, it was necessary to build a computational model. The kinetic modeling software (Kintek GKE (Johnson et al. 2009)) uses an ordinary differential equation (ODE) approach to simulate and fit reaction data. Although this approach provided a convenient method to estimating rate constants, only a limited number of reaction steps could be explicitly modeled, such that the filaments were limited to a size of 4-5 DBD. Since the SgrAI filaments can grow indefinitely and from either end, the limitation of 4-5 DBD per filament in the model required validation before it could confidently be employed in this case. We began by using an "approach to equilibrium" method and the experimental arrangement shown in Figure 14(B) (Park et al., 2018). DNA containing a Flo fluorophore (Flo-40-1 or Flo-PC) was mixed with Rox labeled DNA (Rox-40-1 or Rox-40) in the presence of excess SgrAI, and the gain in emission from Rox with excitation of Flo (a.k.a. the FRET signal) was measured as a function of time. The measurements were repeated in the presence of Ca^{2+} to determine how the presence of a divalent cation affects the association and dissociation kinetics. The data were fit to a simple scheme, shown in Figure 15(C), but with some variations including: 1) filaments assemble and disassemble only *via* individual DBD at either end; 2) filaments were allowed to grow by the association of 2 smaller filaments, and conversely dissociation was allowed to occur *via* disruption between any adjacent DBD in a filament to product two smaller filaments; 3) the association and dissociation rate

constants were allowed to be fit individually for DBD with cleaved and with uncleaved DNA. Since modeling required the prediction of the FRET signal from assemblies of SgrAI bound to Flo and Rox labeled DNAs, the three-dimensional filament structure described above was used to calculate expected distances between fluorophores in the filament to weight the expected FRET signal within filaments with different compositions and organizations. A total of 15 different reaction data sets were used in the global data fitting. The results showed that all variations of the model resulted in the same rate constants (within error), with two exceptions: if filaments dissociate one DBD at a time from the filament ends only, then the dissociation rate constant that fits the experimental data is higher than if filaments dissociated from anywhere along their length. In addition, the association rate constant of DBD was ~10 fold slower in the presence of Ca^{2+} than in its absence. However, given that all models performed similarly in terms of their goodness of fit, the simplest version of the model was used going forward: 1) filaments can be modeled as assembling/disassembling only from their ends or *via* breaking and joining of filaments (Questions #6-7), but the slower rate constant should be used in the latter case; 2) filament assembly and disassembly have the same kinetics whether the DNA is cleaved or not (Question #8); 3) the data fit well (Figure 15(D)) with the same rate constants to models that include filament lengths containing up to 4 DBD and up to 5 DBD, suggesting that modeling of longer filaments is not necessary (i.e. the addition of filaments up to 5 DBD did not significantly improve the fit of the data compared to models including filaments up to 4 DBD). The association rate constant of DBD into filaments was found to be relatively slow (10^5-10^6 $M^{-1}s^{-1}$) which is 3-4 orders of magnitude slower than diffusion limited (Alberty and Hammes 1958; Eigen and Hammes 1963), and the dissociation of DBD from filaments was characterized by rate constants of 0.02-0.1 s^{-1}, depending on the model used.

Our models utilized the knowledge from the FRET titration experiment (Figure 15(A,B)), which suggested that DBD bind filaments with the same kinetics irrespective of filament size. The agreement of the model (shown as solid lines, Figure 15(D)) to the experimental data (points in Figure 15(D)) suggests that this is an adequate initial assumption. However, we tested this assumption further by introducing into the model a cooperativity factor, whereby the dissociation rate constant of a DBD from filaments would depend on the filament size. Specifically, DBD dissociation from shorter filaments would be presumed to be faster than DBD dissociation from longer filaments. To test this idea, we first performed an experiment that assumed no cooperativity. We simulated a distribution of filaments containing up to 12 DBD and compared the modeling results to the distribution of particles measured experimentally using negative stain EM (Figure 15(E)). The simulation of filament lengths (purple, Figure 15(E)) shows excellent agreement with the experimental data (blue, Figure 15(E)). Next, we introduced the cooperativity factor. When we reran the simulation, a very different distribution of filament sizes was observed (red and four shades of green, Figure 15(F)). Notably, once the filament sizes were >3 DBD, the simulations with cooperativity introduced no longer compared well with the experimental negatively stain EM data. Hence, a model with no cooperativity and in which the affinity of DBD for filaments does not change with filament length optimally fits the data. These results also agree with the FRET data (Figure 15(A,B)) and provide additional evidence to address Question #5 in Table 1. Similar simulations also addressed Question #1; the distribution of filament lengths is controlled by the concentration of DNA bound SgrAI, as well as the ratio of the forward and reverse rate constants for assembly of DBD into filaments (Park et al., 2018).

With the validated model for filament assembly developed above, and initial estimates of rate constants for filament assembly and disassembly, the model was next expanded to include DNA cleavage as well as the dissociation of cleaved DNA from non-filament associated DBD. Because the reporter DNA that is used in the cleavage reactions is short, 18 bp, FRET titrations were also performed with 18-1 and activator PC DNA in the presence of excess SgrAI. These experiments yielded a $K_{1/2}$ of 0.5 ± 0.2 μM and Hill coefficient of 2.5 ± 0.3 (Figure 16(A,B)) (Park et al., 2018). A control assay was performed without SgrAI, which showed no change in FRET signal (black, Figure 16(B)). Unlike the FRET titration with 40 bp DNAs (Figure 15(A,B)), the Hill coefficient in the case of the 18 bp DNA shows cooperativity of filament assembly on the longer DNA. This may be due to the weaker binding of ES_{18} (SgrAI bound to 18-1) to filaments as a result of the shorter flanking DNA in the 18 bp DNA (compared to the 40 bp DNA). However, subsequent tests showed no effect of including cooperativity in the computational models on the quality of its fit to the experimental data (Park et al., 2018). In addition, A FRET titration was also performed with Flo-18-1 and Rox-18-1 in the presence of excess SgrAI (Park et al., 2018). As anticipated, the association of DBD with shorter flanking base pairs shows a far weaker association with a $K_{1/2}$ > 400 μM (Figure 16(C-D)). For this

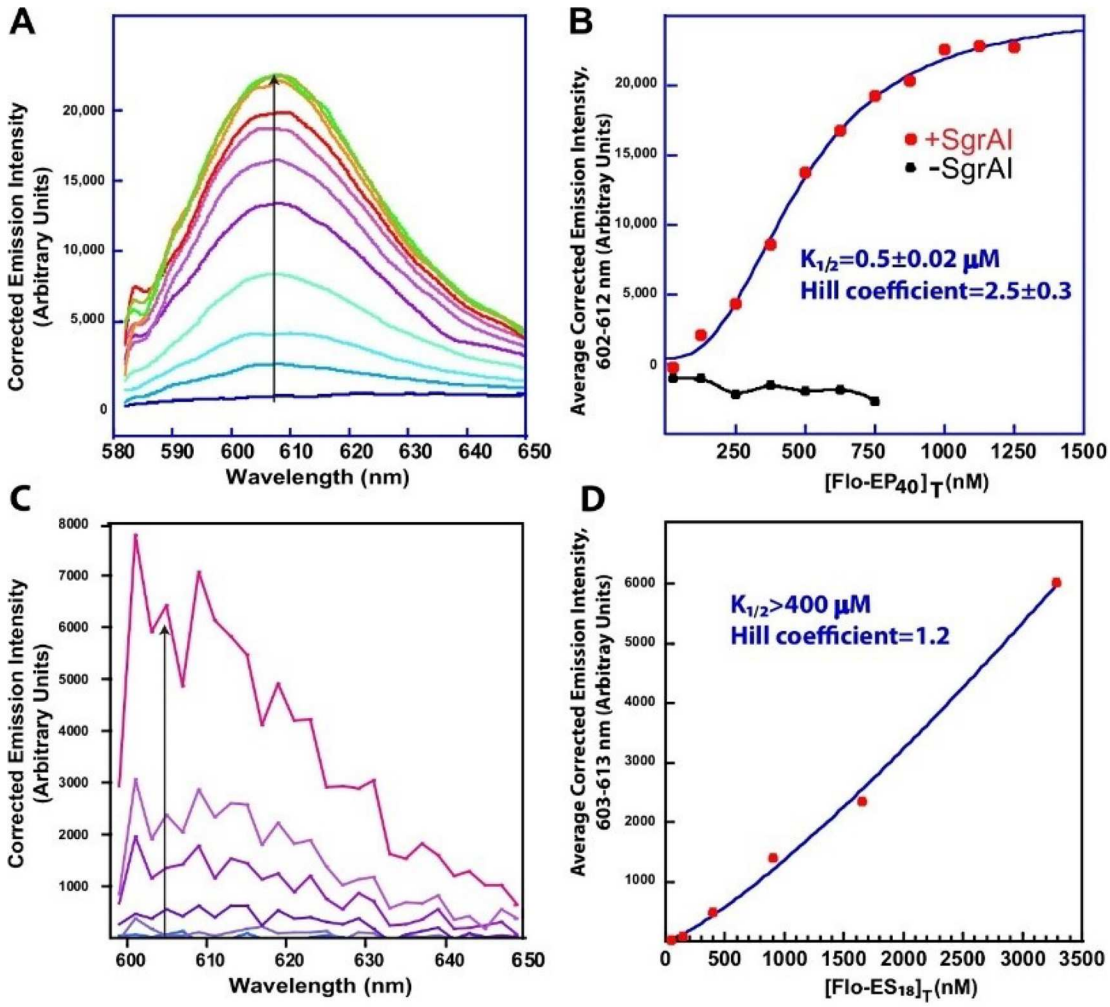

Figure 16. Simplified kinetic model and FRET titrations of ES_{18} with EP_{40}. A. FRET emission (emission of Rox with Flo excitation) of Rox-18-1 with added Flo-PC in the presence of excess SgrAI. The arrow indicates the increasing FRET signal with increasing concentrations of added Flo-PC. B. Fit to data in B showing a $K_{1/2}$ of 0.5 ± 0.2 µM and Hill coefficient of 2.5 ± 0.3 (red points are experimental data, line is fit to Hill equation). The black line and points are from the experiment repeated without SgrAI and shows no change in FRET signal. C. As in B but with Flo-18-1 and Rox-18-1 in the presence of excess SgrAI. D. Data (red circles) from C and attempted fit to the Hill equation. The line shown has a $K_{1/2}$ of over 400 µM and Hill coefficient of 1.2. (Adapted from Park et al. (Park et al., 2018)).

reason, our computational reaction models did not include the assembly of ES_{18} (SgrAI bound to 18-1) or EP_{18} (SgrAI bound to cleaved 18-1) with itself in filaments.

Next, we repeated the approach to equilibrium measurements but using Rox-18-1 with Flo-PC and in the presence of excess SgrAI (Park et al., 2018). The resulting complexes, ES_{18} and EP_{40}, were found to associate with nearly the same association rate constant as EP_{40} and ES_{40}. However, the dissociation rate constant for ES_{18} (0.08 s^{-1}) was much faster than the dissociation rate constant for EP_{40} (0.02 s^{-1}). A faster dissociation rate constant is consistent with the weakened affinity of ES_{18} for EP_{40} suggested by the FRET titrations (Figure 16(A,B)). Rate constants determined by the approach to equilibrium method were used as starting points in global fitting of the DNA cleavage data and constrained to values that fell within their 95% confidence interval.

Figure 17 shows a simplified version of the reaction model using in computational modeling. The activator DNA (PC DNA, brown, Figure 17) first assembles into a 40 bp semi-continuous duplex using its 5′-CCGG-3′ overhangs, and with the rate constants k_1 and k_{-1}, which then binds to SgrAI with rate constants k_2 and k_{-2}. The reporter DNA is shorter, only 18 bp in length (yellow represents uncleaved, purple represents cleaved, Figure 17) and binds to SgrAI with rate constants k_3 and k_{-3}. Filaments assemble with rate constants k_5 and k_{-5}. DNA cleavage within the filament is governed by an irreversible forward step k_6. DNA is also cleaved slowly in the non-filamentous state (not

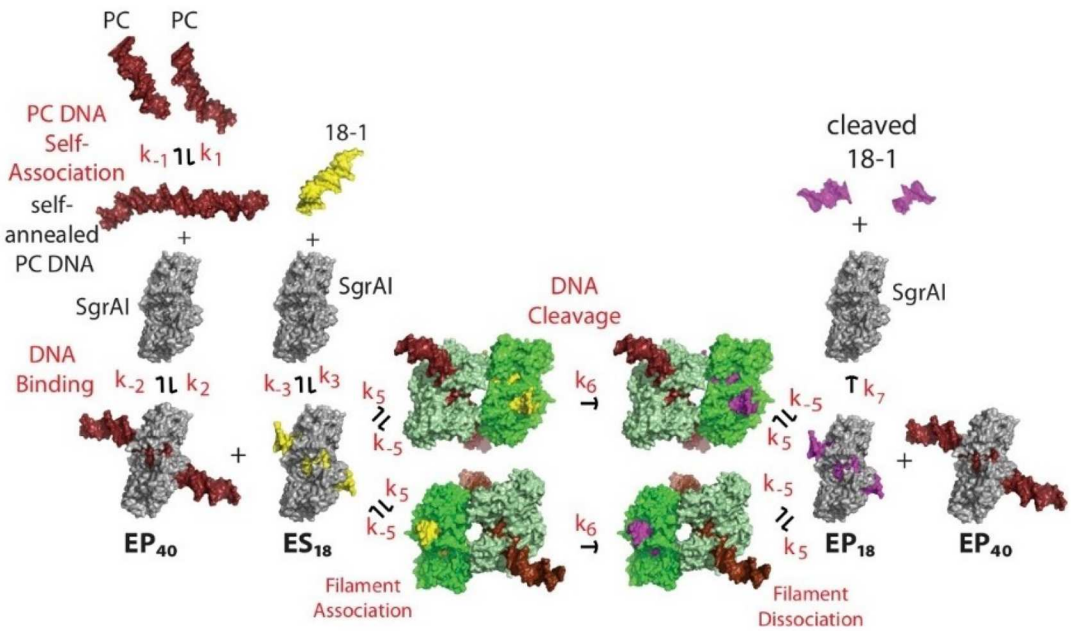

Figure 17. Simplified version of the model used in computational modeling, global data fitting, and simulations. PC, pre-cleaved 40 base pair activator DNA containing a primary recognition sequence, 18-1, 18 bp reporter DNA containing a primary recognition sequence, EP_{40}, SgrAI bound to two copies of self-annealed PC, ES_{18}, SgrAI bound to 18-1, EP_{18}, SgrAI bound to cleaved 18-1. (Adapted from Park et al. (Park et al., 2018)).

shown in Figure 17, but included in the computational model). Cleaved 18-1 is modeled as an irreversible dissociation from SgrAI due to its short length (7 bp) and the reaction temperature (25 °C) and is governed by the rate constant k_7.

Figure 18 shows representative data collected with the three methods described in Figure 14 (Park et al., 2018). The different colors refer to different concentrations of added activator PC DNA, which is expressed as [EP_{40}]. Again, several variations of the mechanistic model were tested for their ability to adequately simulate the experimental data. Variations include maximum filament sizes of 2, 3, and 4 DBD, concerted DNA cleavage within the filament, and assembly/disassembly at both ends of filaments or alternatively also within filaments. The forward rate constants for the self-assembly of PC-DNA and the binding of SgrAI to each type of DNA (18-1 or self-annealed PC DNA) were assumed to be diffusion limited (10^9 $M^{-1}s^{-1}$) (Alberty and Hammes 1958; Eigen and Hammes 1963). The reverse (dissociation) rate constants were constrained by either calculated (in the case of PC DNA self-annealing) or measured K_D (Park et al. 2010, 2018). A total of 22 data sets were used in the global data fitting. Qualities of fit were best for pathways that included 4 DBD but did not suffer significantly in quality of fit when only 2 DBD were used. Follow-up simulations with the model and fitted rate constants indicated that filaments larger than 4 were rare due to the low concentrations of PC DNA and rapid turnover of cleaved DNA under the reaction conditions (Figure 19(A)). Similarly, the inclusion of cooperativity in ES_{18} or EP_{18} binding to EP_{40}, as seen in the FRET titration, was tested by decreasing the dissociation rate constant of the ES_{18} by different degrees. However, no improvement in the fits was observed (Park et al., 2018). Hence, cooperativity in ES_{18} binding to the filaments was not included in the final computational model.

The final rate constants from global fitting are shown in Table 2. All model variations produced similar rate constants. The values shown in parentheses indicate the range of values giving 0.95 times the best χ^2 (a measure of the quality of fit). The rate constant for DNA cleavage in the filament showed a very wide range of values (0.3-6 s^{-1}) hence the values giving fits within 0.99 times the best χ^2 are also given. With these values in hand, we turn back to the questions originally driving this analysis. First, all of our models assumed that the DNA dissociated only from SgrAI DBDs that are not part of a filament. This assumption was based on analysis of the filament structure, which shows: (i) the DNA binding cleft opens toward the center of the filament which is secluded from solvent; (ii) there are numerous contacts between the flanking DNA of one DBD and the neighboring DBD which stabilize the filament (summarized in Figure 11). Dissociation and reassociation of DNA from a stable filament would require a large conformational change

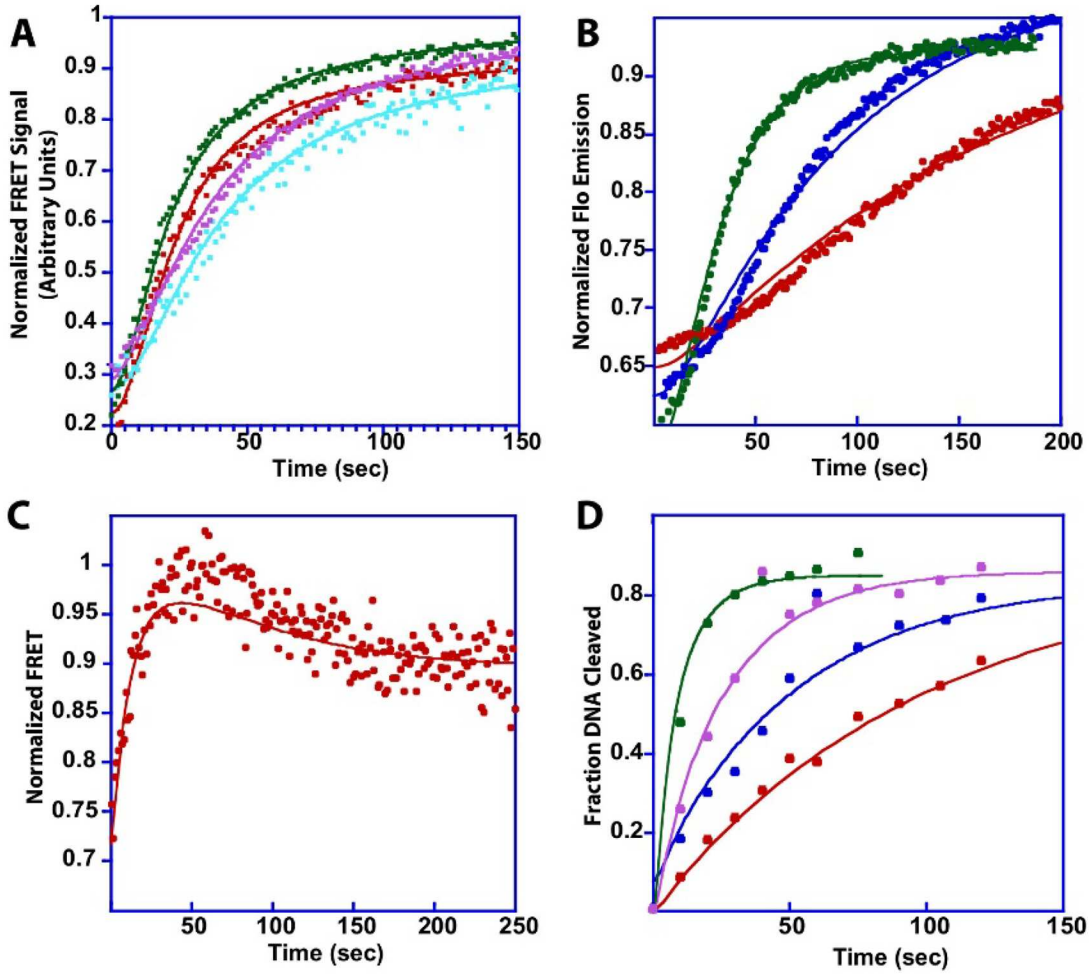

Figure 18. Experimental data used in fitting the computational model with corresponding simulations. A. Comparison of experimental data (points) to computational simulations (lines) for the approach to equilibrium experiment with SgrAI bound to 18-1 and SgrAI bound to PC DNA. B. As in B but for the experiments utilizing Flo-18-1-Rox. C. As in A, but with the data utilizing Flo-PC and Rox-18-1. D. As in a but for the ^{32}P-18-1 cleavage date analyzed by PAGE and autoradiography. (Adapted from Park et al. (Park et al., 2018)).

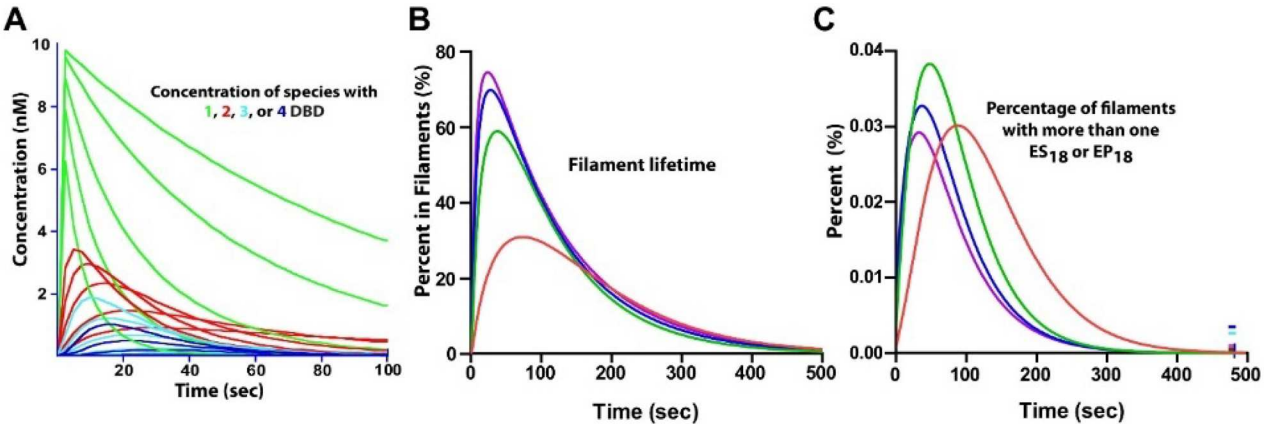

Figure 19. Simulations using the computational kinetic model of the activated DNA cleavage pathway of SgrAI. A. The concentration in nM of species containing reporter DNA and 1, 2, 3, or 4 DBD as a function of time in the experiments used in the global data fitting. The different lines of each color represent reactions with different concentrations of activator DNA used the DNA cleavage measurements used in data fitting. The simulations show that most species are much smaller than 4 DBD. B. As in A, showing the concentrations of filaments containing either 18-1 (in either uncleaved or cleaved states). The plot shows that the residence time of 18-1 in filaments is short-lived due to rapid cleavage, dissociation of filaments, and irreversible dissociation of cleaved 18-1 from SgrAI. C. As in a showing how the filaments with more than one copy of 18-1 varies with the concentration of EP_{40}. Very few such filaments (<0.1%) are found. (Adapted from Park et al. (Park et al., 2018)).

Table 2. Final best rate constants from global data fitting.

Reaction Step	Forward Rate Constant (Range giving 0.95[a] the best fit χ^2)	Reverse Rate Constant
Self-annealing of PC DNA into pseudo-continuous 40-1	2×10^{7} $M^{-1}s^{-1}$ (10^{6} to 6×10^{-8} $M^{-1}s^{-1}$)	8 s^{-1} (0.4 to 225 s^{-1})
SgrAI binding to self-annealed PC DNA to create EP_{40}	10^{9} $M^{-1}s^{-1}$ (fixed)	0.06 s^{-1} (fixed)
SgrAI binding to uncleaved 18-1 to create ES_{18}	10^{9} $M^{-1}s^{-1}$ (fixed)	0.6 s^{-1} (fixed)
Association of EP_{40} with EP_{40} alone or at filament ends	3×10^{5} $M^{-1}s^{-1}$ (1.1×10^{5} to 7×10^{5} $M^{-1}s^{-1}$)	0.017 s^{-1} (0.007 to 0.02 s^{-1})
Association of ES_{18} or EP_{18} with EP_{40} alone or at filament ends	2×10^{5} $M^{-1}s^{-1}$ (2×10^{5} to 3×10^{5} $M^{-1}s^{-1}$)	0.017 s^{-1} (0.007 to 0.02 s^{-1})
DNA cleavage (ES_{18} → EP_{18})	0.8 s^{-1} (0.3-6 s^{-1})(0.4-1.0 s^{-1})[a]	0 (fixed)
Dissociation of cleaved 18-1 from SgrAI	>0.4 s^{-1}	0 (fixed)
DNA cleavage within nonfilamentous ES_{18}	9×10^{-4} $M^{-1}s^{-1}$ (8×10^{-5} to 2×10^{-3} $M^{-1}s^{-1}$)	0 (fixed)

[a]Range of values resulting in fits with a χ^2 with 0.99 * best χ^2.

of the SgrAI dimer to open the DNA binding cleft, which would not be possible without steric clashes between adjacent DBD in the filament and would thus be energetically unfavorable. Further, the approach to equilibrium data fit well to a rapid association/dissociation of DBD in and out of filaments, making dissociation of DNA from isolated DBD plausible. The DNA cleavage data also fit well to the mechanistic model where cleaved DNA is released only from dissociated (i.e. non-filamentous) DBD, and the association/dissociation kinetics also accurately predict the filament size distribution observed by negative stain EM. Finally, association of DBD into the filament is rate limiting, as evidenced by the profound effect of the DNA cleavage rate on the concentration of PC DNA. If filaments formed and remained assembled while cleaved DNA dissociated and uncleaved DNA reassociated, this dependence would not be observed. Thus, we conclude that the answer to Question #2 is that DBD must first dissociate from filaments prior to the dissociation of cleaved DNA.

Question #3 asks if DBD (and their bound cleaved DNA) become trapped in filaments after DNA cleavage occurs. A simulation of the lifetime of a reporter DNA within a filament is shown in Figure 19(B). Concentrations used in the simulation mimicked those of the experimental reaction conditions and show that the lifetime of the reporter DNA in the filament is short (Figure 19(B)). The rapid association/dissociation of DBD into and out of the filament allows for cleaved DNA to be dissociated from SgrAI. Hence trapping *per se* does not occur. However, if concentrations are sufficiently high, a steady state with longer DNA (i.e. that which is stably double stranded after cleavage) and SgrAI into and out of the filament will occur such that much of the DNA is contained within filaments (see distribution of different filament lengths, Figure 15(E)).

Question #4 asks how DNA cleavage can occur if DBD are rapidly cycling in and out of filaments. The rate constants obtained from global data fitting (Table 2) show a 10- to 50-fold faster rate constant for DNA cleavage compared to DBD dissociation (0.8 s^{-1} vs. 0.08 s^{-1} or 0.017 s^{-1}, depending on the length of the DNA). Hence, most of the time, the DNA will be cleaved prior to DBD dissociation from a filament. Finally, Question #9 asks whether DNA in filaments is cleaved in a coordinated manner. The data fit equally well to models that included simultaneous DNA cleavage and to models that considered independent DNA cleavage within each DBD in filaments. The reason for the ambiguity may be explained by several factors. First, DNA cleavage is rapid relative to other steps, and therefore the cleavage rate does not strongly influence the observed data (hence the large range in values at the 0.95*χ^2 threshold). Second, very few filaments contain more than one copy of DBD bound to the reporter DNA (i.e. ES_{18}), and therefore there may be insufficient measurements to definitively discriminate between the cooperative vs. the non-cooperative scenarios (Figure 19(C)). However, the simplest explanation is that cleavage within each DBD is independent of the other DBDs within the filament, provided that the SgrAI enzyme resides in the activated state. Of our 11 questions, only two remain, and that is Questions #10 and #11. As for Question #10, association of DBD into the filament is the rate determining step at most of the activator DNA concentrations used in the assays (Park et al., 2018). Its second order rate constant is 4-5 orders of magnitude slower than diffusion limited. Hence most collisions between DBD appear to not result in filament

formation. The existence of a seemingly unnecessarily slow association step in a mechanism meant to activate an enzyme is counterintuitive. We address this issue further below in a section discussing the biological significance of the filament forming mechanism of SgrAI which also addresses Question #11.

Including the secondary DNA cleavage activity in the computational kinetic reaction model

An important and unusual characteristic of SgrAI is its ability to modulate substrate specificity upon activation. The substrate, here, is duplex DNA with a particular nucleotide sequence. Figure 20 shows two types of models to explain this activity. First, in Figure 20(A), DBD are represented as either blue or orange circles corresponding to SgrAI bound to either primary or secondary sites, respectively. DBD bound to the primary DNA sequence readily filament (top line), while DBD bound to the secondary DNA sequence do not filament (middle line). However, when the two sequences are mixed together, DBD bound to primary sites will stimulate filament formation and draw in DBD bound to a secondary site (lower line, Figure 20). This model explains why, in the absence of primary sequences, SgrAI bound to secondary sequences alone cannot readily form filaments nor cleave DNA. The primary DNA sequences serve two roles. First, they induce filamentation and bring in SgrAI bound to secondary sequences. Second, they activate cleavage of the secondary DNA sequences. The model shown in Figure 20(B) proposes a mechanism to further explain this behavior. In this model, nonfilamentous SgrAI bound to DNA (primary or secondary) equilibrates between a low activity L state, which is observed in crystal structures of individual DBDs (e.g. Figure 3), and the high activity H state, which is observed in filamentous structures of SgrAI bound to DNA (e.g. Figures 10 and 12). Only DBD in the H state associate into filaments, and filaments will form with sufficient concentrations of other DBD that are also in the H state. However, the L state is the baseline low-energy state that is favored over the H state in the absence of filamentation. We propose that the base pair substitutions in the two types of secondary sequences further stabilize the L state over the H state. As a consequence, this would diminish the concentration of H state DBD and their filamentation when SgrAI is bound to a secondary sequence. The purpose of this phenomenon will be discussed in the context of SgrAI biology below.

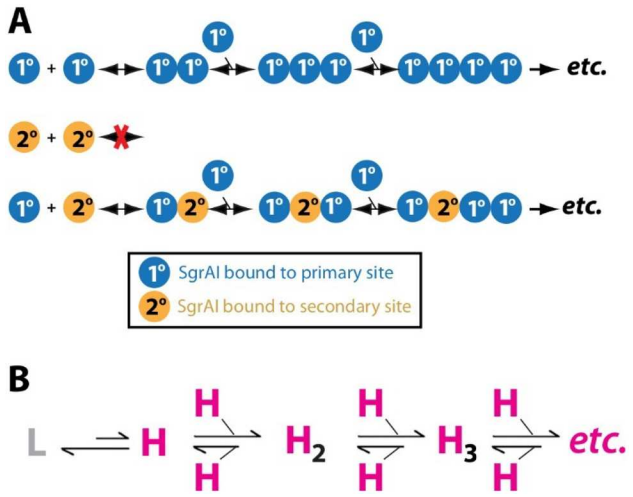

Figure 20. Models of secondary site cleavage by activated, filamentous SgrAI. A. Blue spheres marked with 1° symbolize SgrAI bound to DNA (DBD) containing a primary sequence. Orange spheres marked with 2° symbolize SgrAI bound to DNA containing a secondary sequence. The top line shows how DBD containing the primary sequence readily assemble into filaments, while those containing the secondary sequence do not (second line). However, when both types of DBD are present, DBD containing the primary sequence will stimulate filament formation and draw in DBD bound to secondary sequences (3rd line, lower). B. Model of the control of filamentation by DBD conformation. L, low activity state as exhibited by non-filamentous SgrAI bound to DNA. H, high activity state as exhibited in filamentous SgrAI. SgrAI bound to DNA, either primary or secondary, are in equilibrium between the L and H conformational states. The L state is energetically favored. H states can filament provided the concentration of other H state DBD is sufficiently high. Rapid DNA cleavage occurs in the H state ($0.8\,s^{-1}$), while the L state exhibits only slow DNA cleavage ($9x10^{-4}\ s^{-1}$).

To test this model and quantitate the degree to which binding to secondary sequences shifts the balance between L and H states further toward the L state, we repeated the global data fitting described above to include the cleavage of a secondary sequence. The type A secondary sequence (C**C**CCGGYG) was embedded in the reporter DNA. However, since cleavage of this secondary sequence within an 18 bp construct gave little to no activation in prior assays, a longer reporter DNA was chosen. The construct used, 26 M-2, contains 26 bp with 9 bp flanking either side of the secondary recognition sequence. Flanking DNA of this length was shown to provide nearly the same degree of activation in assays performed with the primary sequence as those with longer flanking DNA (Figure 11(C)). However, the longer reporter DNA required some modification of the computational model described above which assumed the irreversible dissociation of the cleaved reporter DNA. Control experiments showed some reversibility in the dissociation of the cleaved 26 M-2 from SgrAI due its inefficient dissociation into single strands. The model was

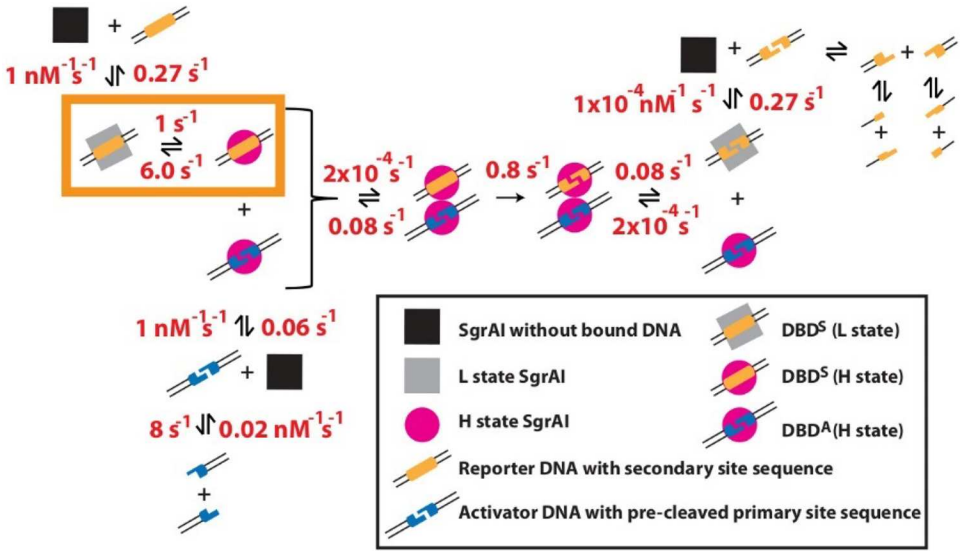

Figure 21. Simplified cartoon of the modified reaction mechanism used in global data fitting of secondary sequence cleavage data with associated rate constants. The orange-yellow boxed region identifies the newly added step to account for the additional stabilization of the low activity state (L) over the high (H) by the secondary sequence substitution. DBD^S, DBD bound to the secondary sequence, DBD^A, DBD bound to the activator DNA. (from Ghadirian, et al. (Ghadirian et al. 2024)).

therefore adjusted to consider this difference (Ghadirian et al. 2024).

To account for the added L state stabilization, an additional equilibrium step was included in the computational model (orange box, Figure 21). This step occurs after DNA binding but before filament assembly, and only for those DBD containing a bound secondary sequence. Although we postulate that this equilibrium occurs for SgrAI bound to both types of recognition sequences, we want to use our computational model calibrated with the cleavage of primary site DNA that did not include the L/H equilibrium step to estimate the additional L state stabilization conferred by the sequence substitutions in the secondary sequence. If our hypothesis is correct, fitting of the kinetic model to the new cleavage data should require only the adjustment of the ratio of the forward and reverse rate constants in the L/H equilibrium. Figure 22 provides examples of the experimental data (points) and corresponding simulation (lines), indicating that the experimental results matched global data fitting from the model. A value of ~6 (with a 95% confidence intervale of 5.4-6.6) was determined for the favorability of the L state over the H state due to the secondary site substitution. Taking into account the effect on DBD affinity to filaments of the longer flanking DNA (i.e. 26 bp vs 18 bp), we reason that this value is closer to 40-fold, corresponding to a 2.2 kcal/mol greater favorability of the L state over the H state as a result of the single base pair substitution in the secondary sequence.

Structural mechanism of filament-induced Expansion of DNA sequence specificity of SgrAI

Our kinetic study of the activated DNA cleavage of secondary sequences by SgrAI hypothesizes that the single base pair change at either the first or second bp of the recognition sequence alters the energy landscape between the low activity (L) and the high activity (H) conformations of SgrAI, such that there is a ~40-fold greater stability of the L conformation. To decipher the origin of this shift in the L/H equilibria, the available structures were analyzed for possible effects of the secondary sequence substitutions. It should be noted that there are no structures of SgrAI in the filamentous state bound to any secondary DNA sequences. The only structures of SgrAI bound to secondary DNA sequences are in the non-filamentous state. Hence, the analyses described below are based on the available structures of SgrAI bound to the primary sequence.

The Type A secondary sequence C**C**CCGGTG (with complementary strand: CACGG**G**G) was used in both the global kinetic analyses and in a crystal structure of SgrAI in the non-filamentous state. The kinetic analyses suggest a 2.2 kcal/mol greater stabilization of the L state over the H state due to the secondary sequence substitution, or that the cost of shifting from the L state to the H is 2.2 kcal/mol greater for the secondary sequence bound SgrAI than primary (Figure 23(A,B)). We can also estimate the difference in L and H state

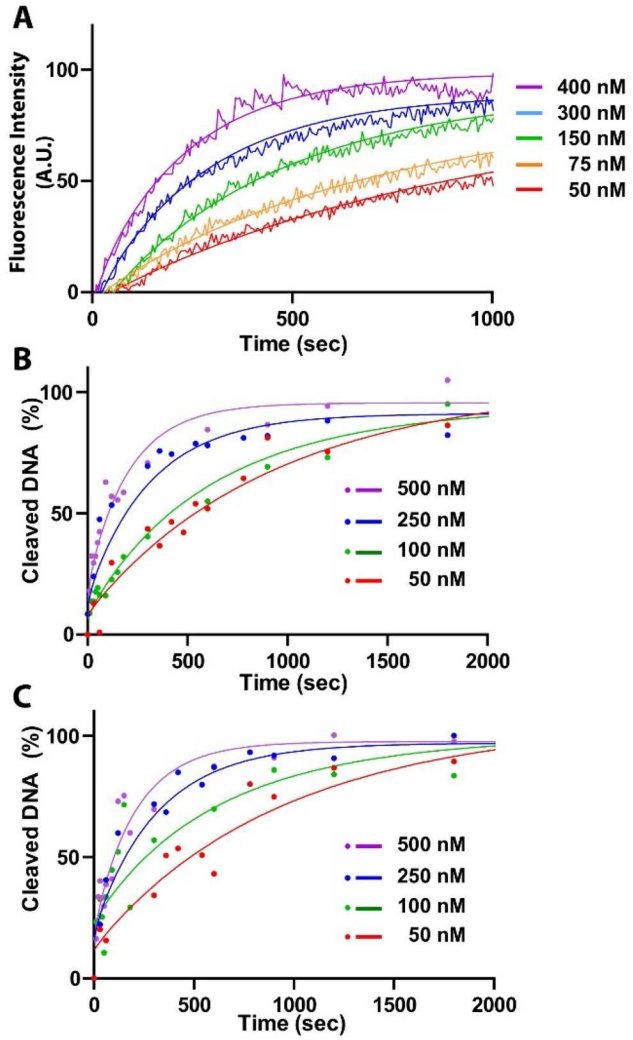

Figure 22. Global data fitting of secondary sequence cleavage by SgrAI. A. Experimental data (jagged lines) and simulations (smooth lines) from DNA cleavage assays and global data fitting, respectively, of fluorescence data using a doubly labeled reporter DNA (Flo-26M-2-Rox) and varied concentrations of PC DNA. B-C. Experimental data (points) and simulations (lines) from DNA cleavage assays utilizing denaturing PAGE and fluorescence quantitation of cleaved reporter DNA. (Adapted from Ghadirian, et al. (Ghadirian et al. 2024)).

stabilities when the primary sequence is bound by comparing the DNA cleavage rate constants for filamentous and non-filamentous primary site bound SgrAI. Assuming that only the H state is active in DNA cleavage, and its maximum cleavage rate constant is $0.8\,s^{-1}$, the cleavage rate constant of non-filamented SgrAI on primary sequences is 9×10^{-4} s^{-1}, which is 900-fold slower, suggesting that the H state is 900-fold less stable than the L state, and corresponding to an energy of 4.0 kcal/mol (Figure 23(B)).

However, modeling this particular secondary sequence into the available structures shows no obvious steric issues or loss of protein-DNA contacts in the L or H state structures. In addition, the structure of SgrAI bound to this secondary sequence in the nonfilamentous state is indistinguishable from the structure of SgrAI bound to a primary sequence (within coordinate error), with the exception of the substituted bases (Little et al. 2011). The 2nd (and 7th) bp of the primary sequence are not recognized by SgrAI using any direct readout mechanisms, that is, there are no hydrogen bonds or van der Waals interactions between SgrAI and the chemically distinct parts of the DNA bases that could explain the observed sequence specificity at these base pairs (Dunten et al. 2008). Instead, the second and seventh base pairs are thought to be specified by indirect readout of the sequence in the form of a DNA distortion which creates a higher energy barrier for non-cognate sequences to attain (Dunten et al. 2008). For example, unstacking of the YR step found between the 7th and 8th nt (TG or CG) is well known to cost little due to its weak stacking interaction (Ulyanov and Zhurkin 1984; Neugebauerová and Kypr 2000; Banerjee et al. 2023). This unstacking is maintained in both L and H states. However, a change in the DNA structure is found between L and H states at the 2nd base step (i.e. between A2 and C3 in the structure). Interestingly, the DNA cleavage site also resides between these two nucleotides. Stacking between A2 and C3 is actually greater in the H state compared to the L state (Figure 23(C,D)). Hence, if this increase in stacking area corresponds to an increase in stacking energy, then these stacking interactions should contribute more favorably to the H state than to the L state. Comparison of the stacking energies of primary sequences at this step (AC and GC) and those of the secondary sequence (CC) shows that the primary sequences result in a greater energy of stacking (Figure 23(E)). Hence, any increase in stacking area is likely to benefit the primary sequence more than the secondary sequence. Accordingly, this would shift the balance between L and H states toward the H state for the primary sequence compared to secondary sequence, and conversely, shift the balance slightly toward the L state for the secondary sequence compared to the primary sequence. Although the H state is intrinsically less stable and higher in energy that the L state for both primary and secondary site sequences bound to SgrAI, the greater DNA stacking makes the H state somewhat less unstable than the L state when the primary sequence is bound compared to when the secondary sequence is bound. Measurements of stacking energy are reported in the literature and vary greatly, but one estimate places the maximum difference in stacking energy between primary and secondary sequences at −1.6 kcal/mol (Banerjee et al. 2023). Meanwhile, the

measurements from DNA cleavage experiments suggest a larger difference of −2.2 kcal/mol (Ghadirian et al. 2024). The difference in these two estimates may be due to the specifics of base stacking, which are structure-dependent but poorly characterized in the literature. Alternatively, the structure of SgrAI bound to a secondary DNA sequence in the filamentous state may reveal yet another heretofore unidentified mechanism for substrate selectivity, which remains to be seen.

Global fitting of kinetic data have not yet been performed using a type B secondary sequence in which substitutions occur in the first base pair. However, single turnover DNA cleavage measurements using radiolabeled DNA and varied concentrations of PC DNA have been performed (40-2B, darkest yellow bars, Figure 6(B)). The results of these cleavage experiments show a similar activation profile as the sequence used in global data fitting described above (40-2A, lightest yellow bars, Figure 6(B)), indicating that a similar perturbation on the L/H equilibrium is likely. Substitutions occur in Type B sequences at the first bp of the recognition sequence where Cyt1 becomes Gua, Ade, or Thy, and Gua8 becomes Cyt, Thy, or Ade. This base pair is recognized by hydrogen bonds from the side chain of Arg31 to Gua8 in both L and H states (Figure 24(A)). Such hydrogen bonds are expected to be absent with a base substitution at Gua8. However, this would be true in both L and H states and is therefore not expected to alter the L/H equilibrium. A model with the Cyt1:Gua8 base pair substituted with Thy1:Ade8 suggests that steric clashes would occur between the backbone at residue 284 in the H state of SgrAI and the C5 methyl group of the Thy1(pink, Figure 24(A–C)); the clashes would not be expected in the L state

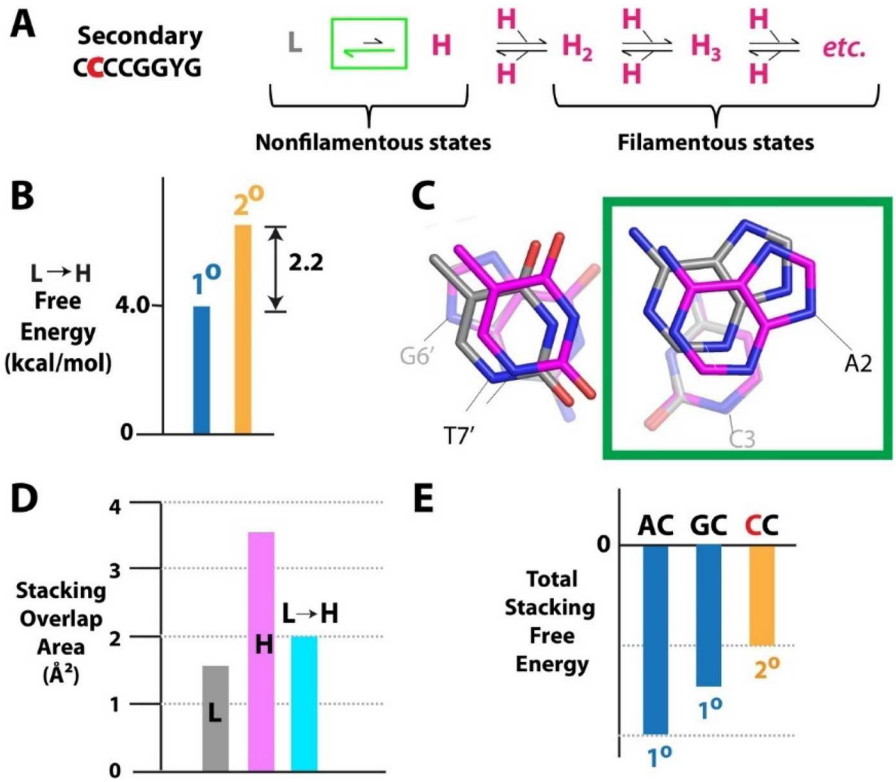

Figure 23. Possible origins of L state stabilization by type a secondary sequence substitutions. A. Hypothesis that the secondary sequence substitution at the second (and seventh) base pair position pushes the L/H equilibrium further to the left, toward the nonfilamenting L state. B. Free energy diagram showing the difference in energy between the H and L states when the secondary sequence (yellow) is bound compared to when the primary (blue) is bound. The double-arrow identifies the difference, estimated as 2.2 kcal/mol from global kinetic modeling, meaning that when the secondary sequence is bound, the L state is 2.2 kcal/mol more favored over the H state compared to when the primary sequence is bound. C. Difference in base stacking at the second base step (between the second and third nucleotide of one strand and the 6th and 7th nucleotide of the recognition sequence on the other). The green boxed region identifies the largest change which is found at the A2-C3 base step in the primary sequence in the L (grey) and H (pink) states. Structures were aligned using all atoms of the C3:G6′ bp (seen in the background). D. Stacking overlap area in Å^2 between the A2 and C3 bases in the L (grey) and H (pink) states. The difference (blue) shows greater stacking area in the H state. E. Comparison of reported stacking free energies in primary site sequences (blue) and secondary (yellow). Primary sequences impart greater energies of stacking, hence the greater stacking area in the H state will result in a greater gain in free energy by SgrAI bound to primary than secondary. (Adapted from polley et al. (Polley et al., 2019)).

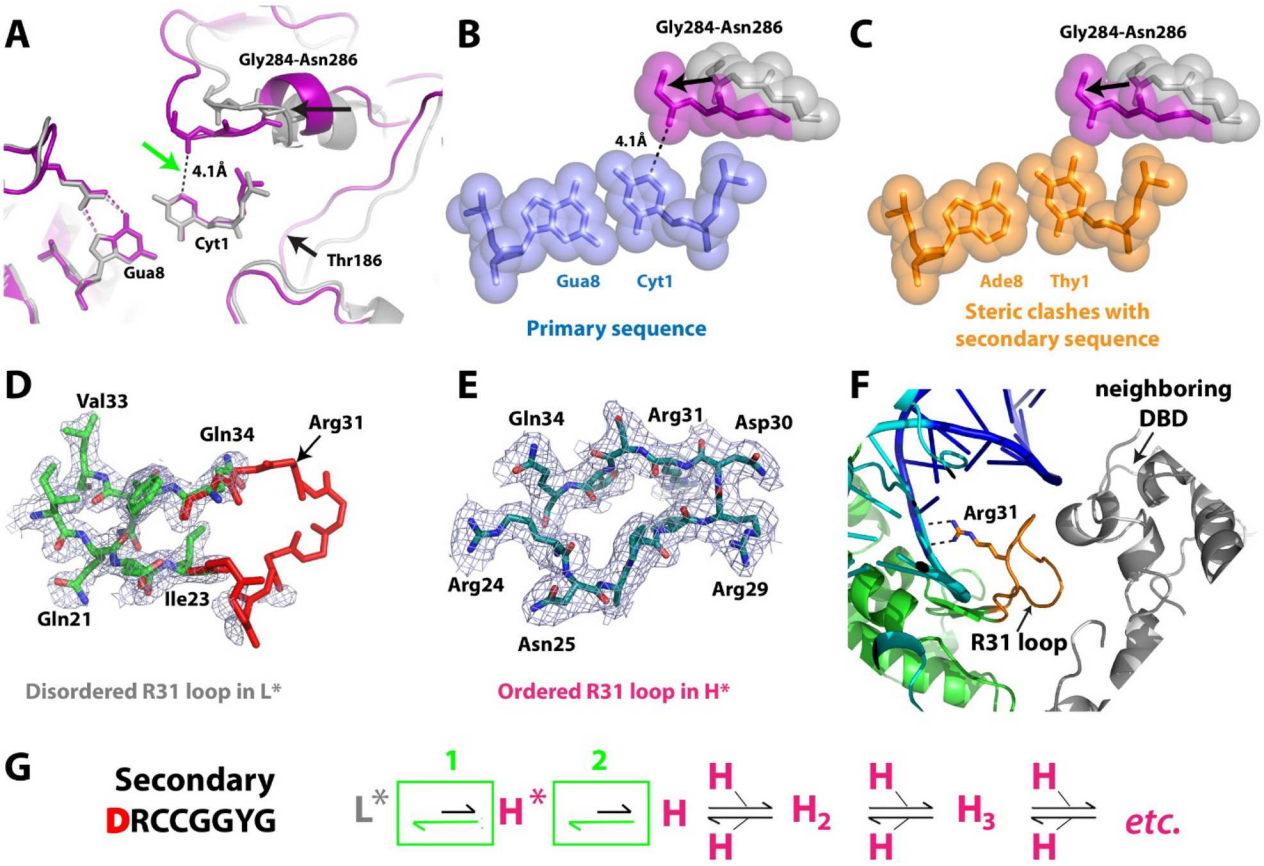

Figure 24. Investigation into possible structural origins of preferential L state stabilization by type B secondary sequences. A. Close approach of the carbonyl oxygen of residue 284 to the base of the first nucleotide (Cyt1) of the recognition sequence in the H state (purple). The L state is shown in grey. B. Close-up of contacts shown in B. C. Substitution of the Cyt1:Gua8 bp with a type B secondary sequence Thy1:Ade8 showing the close contact between residue 284 and the C5 methyl group of Thy1. D. Electron density on the loop containing residue Arg31 (a.k.a. the R31 loop) in the structure of SgrAI without bound DNA showing a break between residues 24-33 (red) likely due to disorder or the multiple conformational states of this loop in the crystal. E. Ordered electron density of the R31 in SgrAI bound to a primary recognition sequence. F. Position of the R31 loop at the protein-protein interface between adjacent copies of SgrAI in the filament. In order to make such contacts, the structure of the loop must be ordered. G. Conformational model of L and H states of SgrAI bound to type B sequences. L* and H* indicate structures in the low and high activity states, respectively, but with the residues of the R31 loop disordered, due to absence of hydrogen bonding between R31 and G8. (Adapted from Shan, et al. (Shan et al. 2022)).

(white, Figure 24(A–C)). Accordingly, these clashes would presumably disfavor the H state and shift the balance to the L state. Substitution of Cyt1 of the primary sequence with Ade or Gua – the other two bases possible in type B secondary sequences – does not show steric overlaps, but the close approach of the backbone carbonyl of residue 284 could potentially block water from hydrogen bonding to the N7 atoms of the purines at position 1 (Shan et al. 2022). Again, high-resolution structures of SgrAI in the filamentous state bound to secondary sequences substituted in the first base pair may be necessary to fully understand the preferential L state stabilization.

Substitution of the Cyt1:Gua8 base pair could result in a second effect. A crystal structure of SgrAI without bound DNA showed a high degree of disorder around Arg31, which is the residue that hydrogen bonds to Gua8 (Figure 24(D)). This segment is well ordered in structures of SgrAI bound to primary sequences (Figure 24(E)). Without the hydrogen bonding interaction between Arg31 and Gua8, the loop containing Arg31 (i.e. the R31 loop, Figure 24(D–F)) may remain in the disordered state even when the Type B secondary sequence is bound. Though ordering of the loop may not be necessary for binding to the secondary sequence, this loop is important for forming protein-protein interactions within the filament (Figure 24(F)). We therefore propose the model shown in Figure 24(G). First, the L* state represents SgrAI bound to the secondary sequence but with a partially disordered R31 loop and in the low activity conformation. The H* state is the H state of this complex but with the R31 loop still disordered. To form filaments, the R31 loop must become ordered to become a true H state. The cost of this disorder-to-order

transition is shown in the equilibrium marked by "2" in Figure 24(G) where the equilibrium favors the H* state over the H state. In addition, the steric conflicts described above destabilize the H* state compared to the L*, shown by the green box and marked "1" in Figure 24(G) where the equilibrium is shifted to the left as a result. The result of these effects is a decrease in the relative concentration of filament-competent H states and reduced filamentation by SgrAI bound to the secondary sequences. These energetic effects however are not so large as to prevent the H state or filamentation as evidenced of the incorporation and activation of SgrAI bound to secondary sequences in filaments stimulated by SgrAI bound to primary sequences.

Biological role of activation and expansion of DNA sequence specificity by filamentation

The above studies establish that SgrAI, a type II restriction endonuclease, cleaves DNA *via* an activated pathway involving helical polymeric filament formation which also expands its sequence specificity from the three cognate primary recognition sequences to seventeen total sequences, including fourteen different secondary recognition sequences. The studies have shown that SgrAI exists in two conformations, a low activity L state, and a high activity H state. The L state is intrinsically more stable, but only the H state is able to assemble filaments. Filaments form with SgrAI bound to primary sequences containing sufficiently long DNA flanks and in the H state, and in turn contacts between SgrAI/DNA complexes within the filament stabilize the H state of the enzyme. The stabilized H state possesses an enhanced DNA cleavage activity, resulting in a ~900-times faster cleavage rate than when in the L state (in the nonfilamentous form). Binding to secondary DNA sequences causes SgrAI to favor the L state by ~40-fold, but when in the H state collisions with other H state SgrAI (such as those formed by SgrAI bound to primary sequences) can result in filamentation and the consequential activation of DNA cleavage.

An important question yet to be addressed in this review is what the biological role of the unusual properties and enzymatic activity of SgrAI are. As a type II RE, SgrAI forms an essential component of the innate immune system of its host organism (*S. cyaneofuscatus*). The enzyme is responsible for cleaving invading phage DNA before its transcription, replication, and perhaps most significantly, methylation by the SgrAI cognate methyltransferase can occur (Wilson and Murray 1991). All REs are essential components of restriction-modification systems, which minimally include the endonuclease and its cognate methyltransferase responsible for methylation (in the case of type II) of the same DNA sequence cleaved by the endonuclease. This system allows the endonuclease to act quickly on invading DNA but not on the host DNA, because the host is protected by methylation. Invading DNA is not expected to be methylated at the SgrAI cleavage sites. However, if the methyltransferase methylates the invading DNA first, then the endonuclease would be unable to cleave and thereby neutralize this DNA.

SgrAI possesses an 8 bp recognition sequence, which is longer than the classic REs such as EcoRI, EcoRV, and BamHI. This may be due to the larger genome size of its host compared to other bacteria (8 Mbp vs. ~4 Mbp). A longer genome means more recognition sites will occur by chance and require methylation to protect from the endonuclease. Methyltransferases use the cofactor SAM, which comes at some expense to the cell. Upregulation of the methyltransferase activity may therefore be too expensive. Accordingly, the longer 8 bp recognition sequence, which occurs more rarely than a shorter sequence, may have evolved to reduce the number of recognition sites in the genome requiring methylation. In addition, the DNA cleavage activity of SgrAI is slow, roughly 80-fold lower than other well characterized REs. However, whereas the combination of slow baseline DNA cleavage rate and the rare recognition sequence would relieve pressure on the host genome, they will also render the enzyme less effective at preventing phage infection. For example, studies have shown that more recognition sites per phage increase the ability of REs to restrict phage infection (Wilson and Murray 1991; Moineau et al. 1993; Kasarjian et al. 2003; Lee et al. 2012). The expansion of DNA sequence specificity to include secondary sequences, as well as the 200 to 1000-fold faster rate of DNA cleavage by activated SgrAI, should greatly improve the effectiveness of SgrAI against invading phage DNA. Because unmethylated primary sequences are required to stimulate filament formation, only the invading DNA will activate SgrAI into forming filaments. SgrAI bound to secondary sequences which occur at much greater frequency than primary can be drawn into filaments stimulated by the phage DNA. Hence many more cleavages will occur on the phage DNA than just at primary sites alone.

Notably, the host genome also contains many secondary sequences that are not methylated by the SgrAI methyltransferase (Figure 25). SgrAI activated by primary sequences on the phage DNA could potentially activate those secondary site-bound SgrAI and cause damage to the host DNA. However, our simulations support the hypothesis that the filament mechanism evolved to ensure rapid neutralization of the invading DNA with

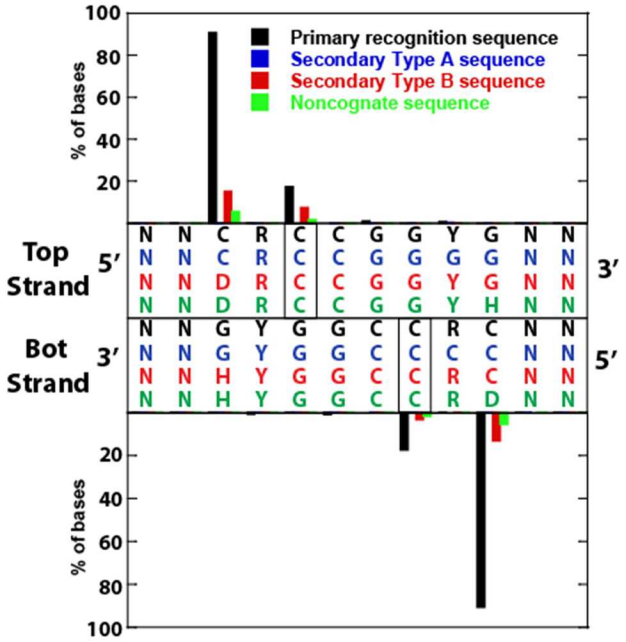

Figure 25. Methylation of SgrAI recognition sequences and a noncognate sequence in the *Streptomyces cyaneofuscatus* genome indicates nearly complete methylation of primary sequences, very little methylation of secondary sequences, and almost no methylation of the noncognate sequence. (Adapted from Ghadirian, et al. (Ghadirian et al. 2024)).

sequestration of its DNA cleavage activity away from the host genome. The key to understanding this effect involves the association kinetics of DNA bound SgrAI into filaments, namely its slow second order rate constant, and the effects of "local concentration" of recognition sites on the phage DNA. For example, to calculate rates of reaction, we need to know both the rate constant and the concentrations of reactants in the cell. A single molecule in an *S. cyaneofuscatus* cell is estimated as 3 nM (see calculations of Table 3). Although we do not know the exact concentration of SgrAI in its host cell, concentrations of enzymes in bacterial cells are known to reach micromolar levels (Albe et al. 1990; Dourado et al. 2021), and it is reasonable to expect that the concentration of SgrAI is in the high nanomolar range (100 copies per cell would give 300 nM). The rate of assembly of two molecules, such as SgrAI bound to sites on separate DNA, can be calculated from the rate constant multiplied by the concentration of each molecule, or $(3x10^5\ M^{-1}s^{-1})(3\times10^{-9}\ M)(3\times10^{-9}\ M)=1.7\times10^{-13}$ M/s. SgrAI bound to secondary sequences sample the H state 40-fold less often than SgrAI bound to primary sequences. Therefore, to calculate the association rate of an SgrAI bound to a primary site on the phage DNA and SgrAI bound to a secondary site in the host genome, we would need to reduce the concentration of the latter by a factor of 40, resulting in a rate of 6.8×10^{-5} M/s.

Table 3. Calculations for determining the concentrations of DNA bound SgrAI.

Concentration of 1 Molecule in a Cell	
Cell volume $CV=\left(\frac{4}{3}\right)\pi r^3$, $r=0.5\,\mu m$	5.24×10^{-16} L
Concentration of 1 molecule in cell	3 nM
$C=\left(\left(\frac{1}{N}\right)/CV\right)$	
N = Avagadro's number (6.02×10^{23} molecules/mole)	
Concentration of 1 molecule in cell if DNA is secondary $=C/40$	0.075 nM
Local Concentrations within contiguous DNA	
Average distance between all sites, primary and secondary, in random sequence DNA (i.e. the distance between a primary site and its two nearest secondary sequences) based on probability	2979 bp
Linear distance between sites along the DNA $D=BP\times3.4\,Å/bp$, BP=distance in bp	10,128 Å
Radius of gyration	1300 Å
$R_g=\left(\frac{(2\times P\times D)}{6}\right)^{1/2}$, P = persistence length of DNA, 500 Å	
Volume occupied by two DNA bound SgrAI $V_g=\left(\frac{4}{3}\right)\pi\left(R_g\right)^3$	9.2×10^{-18} L
Concentration of 1 molecule in Vg	180 nM
$LC=\left(\left(\frac{1}{N}\right)/V_g\right)$	
N = Avagadro's number (6.02×10^{23} molecules/mole)	
Adjusted concentration for H states of secondary site bound SgrAI = LC/40	4.5 nM

Recognition sites on the same DNA molecule, as in phage DNA, experience a higher local concentration relative to each other because they are tethered. The local concentration effect can be calculated by estimating the volume that the two sites occupy. The average distance between recognition sites (primary and secondary) in random sequence DNA is approximately 3 kb (Table 3). From this linear distance along DNA, a volume can be estimated using the calculated radius of gyration of DNA (Table 3). Each SgrAI bound recognition site connected by 3 kb is therefore estimated to appear to have a concentration of 180 nM relative to the other. Again, the concentration of secondary sequence-bound SgrAI should be reduced by a factor of 40. Thus, the assembly rate of SgrAI bound to a primary sequence, and one bound to a secondary sequence can be calculated as $(3x10^5\ M^{-1}s^{-1})(180\times10^{-9}\ M)(4.5\times10^{-9}\ M)=2.4\times10^{-10}$ M/s. The result reveals a rate that is 36,000 times faster than the rate for primary site bound SgrAI sequestered on the phage and

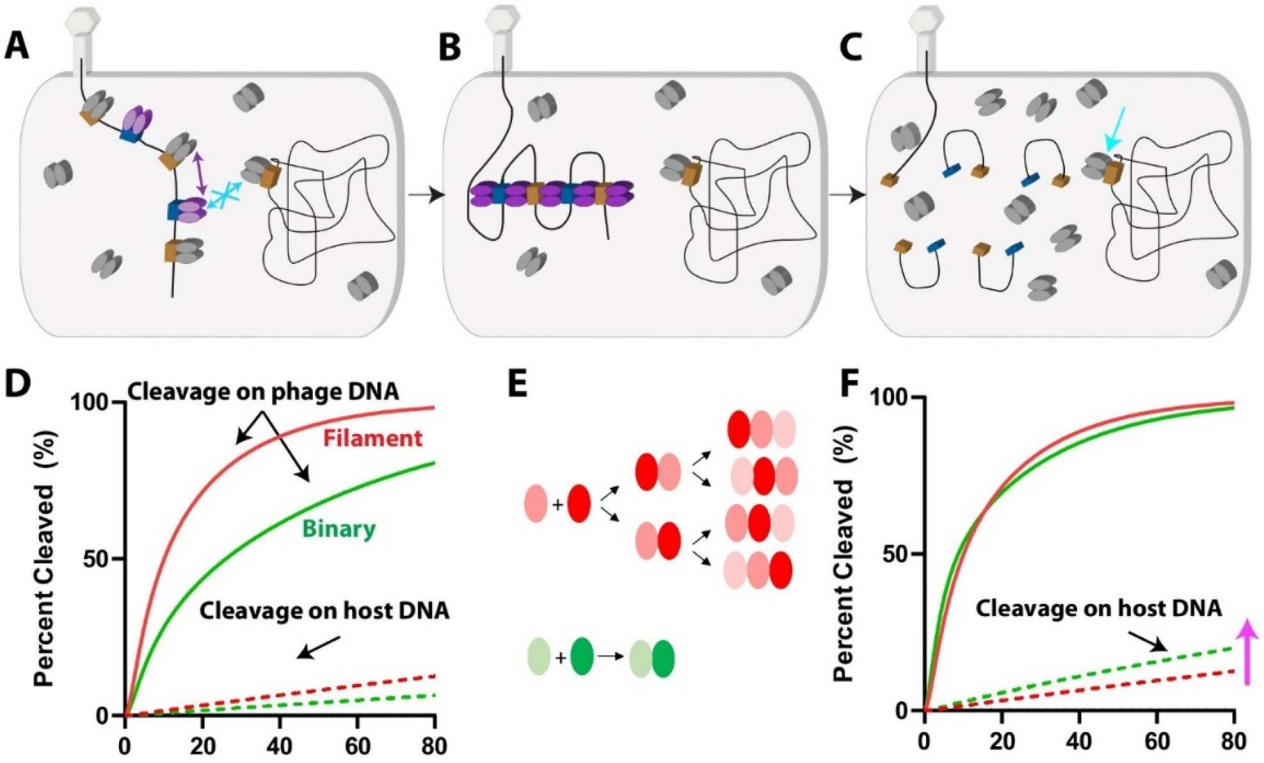

Figure 26. Simulations reveal a role for filamentation in rapid cleavage of invading DNA with minimal damage to the host. A. Cartoon of SgrAI (ovals) in a bacterial cell. SgrAI will bind available secondary sequences (orange rectangles) but without cleaving. Invading DNA contains unmethylated primary (blue rectangles) and secondary sequences which will bind to SgrAI. Primary sequences induce a shift from the low activity conformation (grey ovals) to a high activity conformation (purple ovals) which have a propensity to filament. The slow second order rate constant for association of DNA bound SgrAI into filaments will strongly favor SgrAI bound to the invading DNA (purple double arrow) and exclude host genome bound SgrAI (blue double arrow). B. SgrAI activated by filament formation rapidly cleave the invading DNA. C. The invading DNA is neutralized and the host DNA is left undamaged (blue arrow). D. Simulation of DNA cleavage in the cell. Solid lines: cleavage on invading DNA *via* the filament (red) or a binary assembly (green) mechanism. Dashed lines: cleavage on the host DNA *via* the filament (red) or binary assembly (green) mechanism. The filament mechanism cleaves the invading DNA faster. Both mechanisms show greater cleavage on the invading DNA than the host. E. Schematic illustrating the greater number of ways species can assemble in a filament (red, pink), vs. a binary (green) mechanism. F. As in D, however the association rate constant for filament assembly of the binary mechanism was increased to reach the same rate of DNA cleavage on the invading DNA. Greater damage to the host DNA is predicted (pink arrow).

the secondary site bound SgrAI sequestered on the host genome.

Figure 26(A) illustrates this concept. In the left-most panel, excess SgrAI in the cell binds to secondary sequences (orange rectangles) on the host genome, but genomic primary sites are methylated and therefore not bound by SgrAI. Consequently, filaments will not be induced and the genome bound SgrAI will remain predominantly in the L state (grey ovals) with the bound DNA uncleaved. When phage DNA invades the cells, its primary sequences (blue rectangles) and secondary sequences (orange rectangles) will become bound by SgrAI. The primary site bound SgrAI will flicker into the H state (purple ovals) more frequently than secondary sequence bound SgrAI, but when H state SgrAI collide the filament will assemble (purple double arrow, Figure 26(A)). Filaments on the phage DNA will draw in SgrAI bound to secondary sequences on the phage DNA 36,000 times faster than those on the host genome due to local concentration effects (Figure 26(B)). As a result, filament assembly, activation, and DNA cleavage will occur predominantly on the phage DNA (Figure 26(C)).

To demonstrate this effect further, simulations were carried out to mimic DNA cleavage in the cell, and specifically the cleavage on the phage DNA vs. the host genome. We repeated the simulations published previously (Barahona et al. 2019) with the new knowledge that H states of secondary site bound SgrAI are 40-fold lower in concentration than H states of primary site bound SgrAI (Ghadirian et al. 2024). To simulate cleavage of secondary sites on the phage DNA, the concentrations of 180 nM for primary and 4.5 nM for secondary were used. To simulate cleavage of secondary sequence on the host DNA, concentrations of 3 nM for SgrAI bound to primary sequences on the phage, and 0.075 nM for SgrAI bound to secondary sequences

on the host genome were used. The results are shown in red in Figure 26(D). As predicted, cleavage on the phage DNA occurs far more rapidly than on the host DNA. Hence the filamentation mechanism has successfully sequestered the activated DNA cleavage of secondary sequences away from the host genome. The closer proximity of the recognition sequences on phage lead to faster rates of cleavage. However, this could in principle also be accomplished by a discrete binary multimeric complex rather than a complex that can actively assemble bidirectional filaments of indefinite length. A binary complex is a closed ended species such as two copies of the DNA bound SgrAI dimer assembling into a tetramer. We simulated DNA cleavage on phage and host DNA using a binary mechanism (green, Figure 25(D)) and the same rate constants and concentrations used for the filament mechanism. The binary mechanism also sequestered DNA cleavage on to the phage DNA and away from the host DNA (compare solid to dotted green line, Figure 25(D)), but cleavage on the phage DNA is clearly much slower than in the filamenting mechanism. The faster rate of DNA cleavage performed by the filamenting mechanism is due to the multiple ways in which filaments can assemble (Figure 25(E)), and the absence of competition between primary-primary and primary-secondary type complexes (Barahona et al. 2019). We next considered that the binary reaction could be accelerated if the association rate constant were faster (an increase of 3.5-fold is necessary to reach the same rate of cleavage on the phage DNA) (Figure 25(F)), however a greater degree of damage to the host DNA is now found (green dotted line, pink arrow, Figure 25(F)). Hence the filament mechanism may have evolved to both rapidly cleave phage DNA with minimal damage to the host genome.

In conclusion, these simulations suggest that the filamentation of SgrAI may have evolved to solve very specific problems in its biological niche. First, by reducing its activity in the cell to limit damage to the genome and to reduce the stress on the SAM pool and methyltransferase activity, and then to rapidly deploy on invading DNA without damaging the host. We expect an SgrAI filament to be very short lived, unlike other known filamenting enzymes which can form large cellular superstructures and compartments and last for hours.

Conclusion and future perspectives

The discovery of filament formation by SgrAI was completely unexpected. Our investigations began with the observation of unusual DNA cleavage activity in SgrAI, namely the effect of its primary recognition sequence on both its DNA cleavage rate and sequence specificity. A series of biochemical, biophysical, structural, and kinetic studies were then employed resulting in a mechanistic model for this behavior. SgrAI binds to both types of recognition sequences, primary and secondary, with high affinity. Within these complexes, the conformation of SgrAI and the bound DNA equilibrates between two states, a low activity L state and a high activity H state. In the absence of filamentation, the L state is intrinsically more energetically favorable, however only the H state forms filaments. Within the filament, the H state is reinforced and maintained by protein-protein and protein-DNA interactions between neighboring DNA bound SgrAI dimers (DBD). The H state has an up to 1000-fold greater DNA cleavage activity than the L, which has been traced to the binding of a second divalent cation in each active site of SgrAI when in the H state. Our understanding of the mechanism by which filamentation expands the DNA sequence specificity of SgrAI from three primary to include fourteen additional sequences (i.e. secondary sequences) is incomplete, but our model proposes that binding to the secondary sequences stabilizes the L state even further. As a result, SgrAI bound to secondary sequences sample the H state less frequently, cleave DNA much more slowly, and have a greatly decreased propensity for filamentation. The presence of the primary sequence stimulates cleavage of secondary sequences by drawing in SgrAI bound to both types of sequences into filaments. Hence this model explains why the secondary sequences are cleaved in the presence of primary, but not in its absence.

Global fitting of DNA cleavage data allowed for the creation of a computational model of the reaction. As a result, rate constants for all major steps in the pathway, including activated DNA cleavage and filament association and dissociation, were estimated. Further, many fundamental questions about this unusual enzymatic mechanism were answered. Filament association was found to be the rate limiting step in the experimental measurements where recognition sequences were located on independent DNA molecules. The rate constant for the association of two SgrAI (each bound to primary sequences) was found to be three to four orders of magnitude slower than diffusion limited. We propose that this slow rate constant is due to the preponderance of the L state amongst colliding SgrAI/DNA complexes (i.e. DBD) leading to many unproductive collisions that do not result in filament formation. Comparison of the DNA cleavage rate constants for nonfilamented and filamented SgrAI (when bound to a primary sequence) has led to an estimation of a

900-fold greater stability of the L over the H state. Taking this into account, the rate constant for the association of pure H state SgrAI reaches the diffusion limit. Hence the slow association of DBD into filaments appears to derive largely from the preferential L state stabilization. Finally, the biological significance of this slow filament association rate has been suggested by simulations of *in vivo* activity to function to sequester DNA cleavage activity on invading DNA and away from the host genome. Further, comparisons of the filament mechanism to a more discrete oligomer mechanism suggest that the filament mechanism is superior in both speed and sequestration activity.

Despite this large body of work, several unanswered questions remain, such as the origin of the greater stabilization of the L state by bound secondary sequences. Some ideas have emerged from current studies which include the use of sequence specific DNA structural energetics, steric overlaps, and a disorder-to-order transition in a segment of SgrAI. In addition, the role of preferential L state stabilization in slowing the association step of filament formation and sequestering cleavage to a single DNA molecule can be tested by perturbing the energetic balance between L and H states using point mutations in SgrAI. Predictions made by computational simulations in the superiority of the filament mechanism in rapid DNA cleavage with limited host genome damage should also be tested by comparison of filament forming and discrete oligomer-forming enzymes. Finally, single molecule methods and stochastic kinetic modeling will advance the abilities of computational modeling of *in vivo* activity.

Curiously, one of the earliest crystal structures showed two copies of SgrAI bound to a primary sequence containing DNA engaged in a domain swapped dimer using the segments at the amino terminus unique to SgrAI and also important to filament formation (Park et al. 2010). However, subsequent cryo-EM studies have not shown evidence of domain swapping in the filament structures. Hence it remains an unanswered question whether the original observation was an artifact of crystallization or perhaps represents a yet unknown mode of action of SgrAI. Also worth considering is that many type II REs from *Streptomyces* do not possess a filament mechanism (Bilcock et al. 1999). If the filament mechanism of SgrAI evolved due to the special circumstances of the larger genome within *Streptomyces*, how do these other systems solve the same problem? It is worth noting that a only relatively small fraction of known type II restriction endonucleases have been examined mechanistically, and also that very few filament forming enzymes have been studied at the level of detail that SgrAI has been studied.

Filamentous enzymes have now been found in more than 30 distinct enzymes derived from diverse organisms and cell types (Park and Horton 2019). Our studies leading to the discovery of filament formation in SgrAI began from a biochemical observation of its DNA cleavage activity. In other systems, filamentation was discovered by differing means. For example, the discovery of enzyme filamentation over fifty years ago in metabolic enzymes purified from their native sources occurred *via* characterization by size exclusion chromatography, light scattering, ultracentrifugation and/or electron microscopy (Olson and Anfinsen 1952; Vagelos et al. 1962; Parmeggiani et al. 1966; Olsen et al. 1970). More recently, filamenting enzymes were discovered when determining cellular location (Narayanaswamy et al. 2009; Noree et al. 2010; Suresh et al. 2015; Shen et al. 2016; Noree et al. 2019). Some enzymes which form cellular self-assemblies were studied biochemically and found to form polymeric filaments (Zhong et al. 2022; Guo and Liu 2023; Hvorecny and Kollman 2023). On the other hand, it is likely that many enzymes form cellular self-assemblies but do not form polymeric filaments, and vice versa. For example, the formation of cellular self-assemblies large enough to be visible by fluorescence microscopy by SgrAI has not been observed to date. This is likely due to the transient, and limited nature of the polymeric filaments of SgrAI within cells. As for the role of filamentation in enzyme regulation, studies have found that filamentation may activate, inhibit, sequester, alter specificity or cooperativity, or change the response to allosteric effectors or PTM (Park and Horton 2019; Hvorecny and Kollman 2023; Calise et al 2024). Many filamenting enzymes are located at crucial and highly regulated branch points in metabolic pathways (Hvorecny and Kollman 2023). Because only recently has the interest in enzyme filamentation been rekindled, there is no doubt that much is still left to discover within this interesting class of enzymes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Science Foundation under Grant MCB-1934291 (N.C.H. and D.L.); the University of Arizona Research, Innovation & Impact (RII) and Technology Research Initiative Fund/Improving Health and Access and Workforce Development (N.C.H.); the National Science Foundation under Grant MCB 2048095(D.L); and the Hearst Foundations Developmental Chair (D.L.).

References

Aggarwal AK. 1995. Structure and function of restriction endonucleases. Curr Opin Struct Biol. 5(1):11–19. doi:10.1016/0959-440x(95)80004-k.

Albe KR, Butler MH, Wright BE. 1990. Cellular concentrations of enzymes and their substrates. J Theor Biol. 143(2):163–195. doi:10.1016/s0022-5193(05)80266-8.

Alberty R, Hammes G. 1958. Application of the theory of diffusion-controlled reactions to enzyme kinetics. J Phys Chem. 62(2):154–159. doi:10.1021/j150560a005.

Anthony SA, Burrell AL, Johnson MC, Duong-Ly KC, Kuo YM, Simonet JC, Michener P, Andrews A, Kollman JM, Peterson JR. 2017. Reconstituted Impdh polymers accommodate both catalytically active and inactive conformations. Mol Biol Cell. 28(20):2600–2608. doi:10.1091/mbc.E17-04-0263.

Aughey GN, Grice SJ, Shen QJ, Xu Y, Chang CC, Azzam G, Wang PY, Freeman-Mills L, Pai LM, Sung LY, et al. 2014. Nucleotide synthesis is regulated by cytoophidium formation during neurodevelopment and adaptive metabolism. Biol Open. 3(11):1045–1056. doi:10.1242/bio.201410165.

Babic AC, Little EJ, Manohar VM, Bitinaite J, Horton NC. 2008. DNA distortion and specificity in a sequence-specific endonuclease. J Mol Biol. 383(1):186–204. doi:10.1016/j.jmb.2008.08.032.

Banerjee A, Anand M, Kalita S, Ganji M. 2023. Single-molecule analysis of DNA base-stacking energetics using patterned DNA nanostructures. Nat Nanotechnol. 18(12):1474–1482. doi:10.1038/s41565-023-01485-1.

Barahona CJ, Basantes LE, Tompkins KJ, Heitman DM, Chukwu BI, Sanchez J, Sanchez JL, Ghadirian N, Park CK, Horton NC. 2019. The need for speed: run-on oligomer filament formation provides maximum speed with maximum sequestration of activity. J Virol. 93(5):1–19. doi:10.1128/JVI.01647-18.

Beese LS, Steitz TA. 1991. Structural basis for the 3′-5′ exonuclease activity of escherichia coli DNA polymerase I: a two metal ion mechanism. Embo J. 10(1):25–33. doi:10.1002/j.1460-2075.1991.tb07917.x.

Bilcock DT, Daniels LE, Bath AJ, Halford SE. 1999. Reactions of Type II restriction endonucleases with 8-base pair recognition sites. J Biol Chem. 274(51):36379–36386. doi:10.1074/jbc.274.51.36379.

Bitinaite J, Schildkraut I. 2002. Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease. Proc Natl Acad Sci U S A. 99(3):1164–1169. doi:10.1073/pnas.022346799.

Brody RS, Frey PA. 1981. Unambiguous determination of the stereochemistry of nucleotidyl transfer catalyzed by DNA polymerase I from escherichia coli. Biochemistry. 20(5):1245–1252. doi:10.1021/bi00508a030.

Burgers PM, Eckstein F. 1979. Stereochemistry of internucleotide bond formation by polynucleotide phosphorylase from micrococcus luteus. Biochemistry. 18(3):450–454. doi:10.1021/bi00570a010.

Calise SJ, O'Neill AG, Burrell AL, Dickinson MS, Molfino J, Clarke C, Quispe J, Sokolov D, Buey RM, Kollman JM. 2024. Light-Sensitive Phosphorylation Regulates Retinal IMPDH1 Activity and Filament Assembly. J Cell Biol. 223(4):e202310139. doi:10.1083/jcb.202310139.

Cassano AG, Anderson VE, Harris ME. 2004. Understanding the transition states of phosphodiester bond cleavage: Insights from heavy atom isotope effects. Biopolymers. 73(1):110–129. doi:10.1002/bip.10517.

Chevalier B, Sussman D, Otis C, Noël A-J, Turmel M, Lemieux C, Stephens K, Monnat RJ, Stoddard BL. 2004. Metal-dependent DNA cleavage mechanism of the I-crei lagli-dadg homing endonuclease. Biochemistry. 43(44):14015–14026. doi:10.1021/bi048970c.

Christofi E, Barran P. 2023. Ion Mobility Mass Spectrometry (Im-Ms) for structural biology: Insights gained by measuring mass, charge, and collision cross section. Chem Rev. 123(6):2902–2949. doi:10.1021/acs.chemrev.2c00600.

Daniels LE, Wood KM, Scott DJ, Halford SE. 2003. Subunit assembly for DNA cleavage by restriction endonuclease SgrAI. J Mol Biol. 327(3):579–591. doi:10.1016/s0022-2836(03)00143-8.

Deibert M, Grazulis S, Janulaitis A, Siksnys V, Huber R. 1999. Crystal structure of muni restriction endonuclease in complex with cognate DNA at 1.7 a resolution. Embo J. 18(21):5805–5816. doi:10.1093/emboj/18.21.5805.

Deibert M, Grazulis S, Sasnauskas G, Siksnys V, Huber R. 2000. Structure of the tetrameric restriction endonuclease ngomiv in complex with cleaved DNA. Nat Struct Biol. 7(9):792–799. doi:10.1038/79032.

Dietrich HM, Righetto RD, Kumar A, Wietrzynski W, Trischler R, Schuller SK, Wagner J, Schwarz FM, Engel BD, Müller V, et al. 2022. Membrane-Anchored hdcr nanowires drive hydrogen-powered co(2) fixation. Nature. 607(7920):823–830. doi:10.1038/s41586-022-04971-z.

Dourado H, Mori M, Hwa T, Lercher MJ. 2021. On the optimality of the enzyme-substrate relationship in bacteria. PLoS Biol. 19(10):e3001416. doi:10.1371/journal.pbio.3001416.

Dunten P, Horton NC, Little EJ. 2008. SgrAI with cognate DNA and Ca^{2+} Bound.

Dunten PW, Little EJ, Gregory MT, Manohar VM, Dalton M, Hough D, Bitinaite J, Horton NC. 2008. The structure of SgrAI bound to DNA; recognition of an 8 base pair target. Nucleic Acids Res. 36(16):5405–5416. doi:10.1093/nar/gkn510.

Eigen M, Hammes GG. 1963. Elementary steps in enzyme reactions (as studied by relaxation spectrometry). Adv Enzymol Relat Subj Biochem. 25:1–38. doi:10.1002/9780470122709.ch1.

Etzkorn C, Horton NC. 2004. Ca2+ binding in the active site of Hincii: Implications for the catalytic mechanism. Biochemistry. 43(42):13256–13270. doi:10.1021/bi0490082.

Etzkorn C, Horton NC. 2004. Mechanistic insights from the structures of Hincii bound to cognate DNA cleaved from addition of Mg2+ and Mn2+. J Mol Biol. 343(4):833–849. doi:10.1016/j.jmb.2004.08.082.

Ferreira APS, Cassago A, Gonçalves KdA, Dias MM, Adamoski D, Ascenção CFR, Honorato RV, de Oliveira JF, Ferreira IM, Fornezari C, et al. 2013. Active glutaminase C Self-assembles into a supratetrameric oligomer that can be disrupted by an allosteric inhibitor. J Biol Chem. 288(39):28009–28020. doi:10.1074/jbc.M113.501346.

Fuller GG, Han T, Freeberg MA, Moresco JJ, Ghanbari Niaki A, Roach NP, Yates JR, 3rd, Myong S, Kim JK. 2020. RNA promotes phase separation of glycolysis enzymes into yeast g bodies in hypoxia. Elife 9:e48480. doi: 10.7554/eLife.48480.

Gemmen GJ, Millin R, Smith DE. 2006. DNA looping by two-site restriction endonucleases: heterogeneous probability distributions for loop size and unbinding force. Nucleic Acids Res. 34(10):2864–2877. doi:10.1093/nar/gkl382.

Gemmen GJ, Millin R, Smith DE. 2006. Tension-dependent DNA cleavage by restriction endonucleases: two-site enzymes are "Switched Off" at low force. Proc Natl Acad Sci U S A. 103(31):11555–11560. doi:10.1073/pnas.0604463103.

Gerlt JA. 1993. Mechanistic principles of enzyme-catalyzed cleavage of phosphodiester bonds. In: Lloyd S, Linn S, Roberts R, editors. Nucleases. 2nd Ed. Cold Spring Harbor Press. p. 1–34.

Ghadirian N, Morgan RD, Horton NC. 2024. DNA sequence control of enzyme filamentation and activation of the SgrAI endonuclease. Biochemistry. 63(3):326–338. doi:10.1021/acs.biochem.3c00313.

Grazulis S, Deibert M, Rimseliene R, Skirgaila R, Sasnauskas G, Lagunavicius A, Repin V, Urbanke C, Huber R, Siksnys V. 2002. Crystal structure of the Bse634i restriction endonuclease: comparison of two enzymes recognizing the same DNA sequence. Nucleic Acids Res. 30(4):876–885. doi:10.1093/nar/30.4.876.

Guo CJ, Liu JL. 2023. Cytoophidia and filaments: you must unlearn what you have learned. Biochem Soc Trans. 51(3):1245–1256. doi:10.1042/BST20221410.

Hingorani-Varma K, Bitinaite J. 2003. Kinetic analysis of the coordinated interaction of SgrAI restriction endonuclease with different DNA targets. J Biol Chem. 278(41):40392–40399. doi:10.1074/jbc.M304603200.

Horton JR, Cheng X. 2000. Pvuii Endonuclease contains two calcium ions in active sites. J Mol Biol. 300(5):1049–1056. doi:10.1006/jmbi.2000.3938.

Horton NC, Connolly BA, Perona JJ. 2000. Inhibition of ecorv endonuclease by deoxyribo-3′-S-Phosphorothiolates: A high resolution X-ray crystallographic study. J Am Chem Soc. 122(14):3314–3324. doi:10.1021/ja993719j.

Horton NC, Perona JJ. 2001 Making the most of metal ions. [Comment]. Nat Struct Biol. 8(4):290–293. doi:10.1038/86149.

Horton NC, Dorner LF, Perona JJ. 2002b. Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. Nat Struct Biol. 9(1):42–47. doi:10.1038/nsb741.

Horton NC, Newberry KJ, Perona JJ. 1998. Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. Proc Natl Acad Sci U S A. 95(23):13489–13494. doi:10.1073/pnas.95.23.13489.

Horton NC, Otey C, Lusetti S, Sam MD, Kohn J, Martin AM, Ananthnarayan V, Perona JJ. 2002. Electrostatic contributions to site specific DNA cleavage by ecorv endonuclease. Biochemistry. 41(35):10754–10763. doi:10.1021/bi020305l.

Horton NC, Perona JJ. 1998. Recognition of flanking DNA sequences by ecorv endonuclease involves alternative patterns of water-mediated contacts. J Biol Chem. 273(34):21721–21729. doi:10.1074/jbc.273.34.21721.

Horton NC, Perona JJ. 1998. Role of protein-induced bending in the specificity of DNA recognition: Crystal structure of ecorv endonuclease complexed with D(Aaagat) + D(Atctt). J Mol Biol. 277(4):779–787. doi:10.1006/jmbi.1998.1655.

Horton NC, Perona JJ. 2004. DNA cleavage by ecorv endonuclease: two metal ions in three metal ion binding sites. Biochemistry. 43(22):6841–6857. doi:10.1021/bi0499056.

Horton NC. 2008. DNA nucleases. In: Rice PA, Correll CC, editors. Protein-nucleic acid interactions structural biology. Cambridge; The Royal Society of Chemistry. p. 333–363.

Hugener J, Xu J, Wettstein R, Ioannidi L, Velikov D, Wollweber F, Henggeler A, Matos J, Pilhofer M. 2024. Filamentid reveals the composition and function of metabolic enzyme polymers during gametogenesis. Cell. 187(13):3303–3318 e3318. doi:10.1016/j.cell.2024.04.026.

Hunkeler M, Hagmann A, Stuttfeld E, Chami M, Guri Y, Stahlberg H, Maier T. 2018. Structural basis for regulation of human acetyl-coa carboxylase. Nature. 558(7710):470–474. doi:10.1038/s41586-018-0201-4.

Hvorecny KL, Kollman JM. 2023. Greater than the sum of parts: mechanisms of metabolic regulation by enzyme filaments. Curr Opin Struct Biol. 79:102530.

Ibstedt S, Sideri TC, Grant CM, Tamás MJ. 2014. Global analysis of protein aggregation in yeast during physiological conditions and arsenite stress. Biol Open. 3(10):913–923. doi:10.1242/bio.20148938.

Johnson KA, Simpson ZB, Blom T. 2009. Global kinetic explorer: A new computer program for dynamic simulation and fitting of kinetic data. Anal Biochem. 387(1):20–29. doi:10.1016/j.ab.2008.12.024.

Joshi HK, Etzkorn C, Chatwell L, Bitinaite J, Horton NC. 2006. Alteration of sequence specificity of the type II restriction endonuclease hincii through an indirect readout mechanism. J Biol Chem. 281(33):23852–23869. doi:10.1074/jbc.M512339200.

Kamerlin SC, Wilkie J. 2007. The role of metal ions in phosphate ester hydrolysis. Org Biomol Chem. 5(13):2098–2108. doi:10.1039/b701274h.

Kasarjian JK, Iida M, Ryu J. 2003. New restriction enzymes discovered from escherichia coli clinical strains using a plasmid transformation method. Nucleic Acids Res. 31(5):e22–22. doi:10.1093/nar/gng022.

Keppeke GD, Chang CC, Peng M, Chen LY, Lin WC, Pai LM, Andrade LEC, Sung LY, Liu JL. 2018. Imp/Gtp balance modulates cytoophidium assembly and impdh activity. Cell Div. 13(1):5. doi:10.1186/s13008-018-0038-0.

Kim EE, Wyckoff HW. 1991. Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. J Mol Biol. 218(2):449–464. doi:10.1016/0022-2836(91)90724-k.

Lambert AR, Sussman D, Shen B, Maunus R, Nix J, Samuelson J, Xu SY, Stoddard BL. 2008. Structures of the rare-cutting restriction endonuclease noti reveal a unique metal binding fold involved in DNA binding. Structure. 16(4):558–569. doi:10.1016/j.str.2008.01.017.

Lassila JK, Zalatan JG, Herschlag D. 2011. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. Annu Rev Biochem. 80(1):669–702. doi:10.1146/annurev-biochem-060409-092741.

Lee JY, Chang J, Joseph N, Ghirlando R, Rao DN, Yang W. 2005. Muth complexed with hemi- and unmethylated dnas: coupling base recognition and DNA cleavage. Mol Cell. 20(1):155–166. doi:10.1016/j.molcel.2005.08.019.

Lee S, Ward TJ, Siletzky RM, Kathariou S. 2012. Two novel type II restriction-modification systems occupying genomically equivalent locations on the chromosomes of listeria monocytogenes strains. Appl Environ Microbiol. 78(8):2623–2630. doi:10.1128/AEM.07203-11.

Lesser DR, Kurpiewski MR, Jen-Jacobson L. 1990. The energetic basis of specificity in the Eco Ri endonuclease–DNA interaction. Science. 250(4982):776–786. doi:10.1126/science.2237428.

Little EJ, Babic AC, Horton NC. 2008. Early Interrogation and Recognition of DNA Sequence by Indirect Readout. Structure. 16(12):1828–1837. doi:10.1016/j.str.2008.09.009.

Little EJ, Dunten PW, Bitinaite J, Horton NC. 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(Pt 1):67–74. doi:10.1107/S0907444910047785.

Liu JL. 2010. Intracellular compartmentation of Ctp synthase in drosophila. J Genet Genomics. 37(5):281–296. doi:10.1016/S1673-8527(09)60046-1.

Lynch EM, Kollman JM, Webb BA. 2020. Filament formation by metabolic enzymes-a new twist on regulation. Curr Opin Cell Biol. 66:28–33. doi:10.1016/j.ceb.2020.04.006.

Lynch EM, Kollman JM. 2020. Coupled structural transitions enable highly cooperative regulation of human Ctps2 filaments. Nat Struct Mol Biol. 27(1):42–48. doi:10.1038/s41594-019-0352-5.

Lyumkis D, Talley H, Stewart A, Shah S, Park CK, Tama F, Potter CS, Carragher B, Horton NC. 2013. Allosteric regulation of DNA cleavage and sequence-specificity through run-on oligomerization. Structure. 21(10):1848–1858. doi:10.1016/j.str.2013.08.012.

Ma X, Shah S, Zhou M, Park CK, Wysocki VH, Horton NC. 2013. Structural analysis of activated SgrAI -DNA oligomers using ion mobility mass spectrometry. Biochemistry. 52(25):4373–4381. doi:10.1021/bi3013214.

Manakova E, Grazulis S, Zaremba M, Tamulaitiene G, Golovenko D, Siksnys V. 2012. Structural mechanisms of the degenerate sequence recognition by Bse634i restriction endonuclease. Nucleic Acids Res. 40(14):6741–6751. doi:10.1093/nar/gks300.

Martin AM, Horton NC, Lusetti S, Reich NO, Perona JJ. 1999. Divalent metal dependence of site-specific DNA binding by ecorv endonuclease. Biochemistry. 38(26):8430–8439. doi:10.1021/bi9905359.

Mizuuchi K, Nobbs TJ, Halford SE, Adzuma K, Qin J. 1999. A new method for determining the stereochemistry of DNA cleavage reactions: Application to the Sfii and Hpaii restriction endonucleases and to the Mua Transposase. Biochemistry. 38(14):4640–4648. doi:10.1021/bi990054p.

Moineau S, Pandian S, Klaenhammer TR. 1993. Restriction/modification systems and restriction endonucleases are more effective on lactococcal bacteriophages that have emerged recently in the dairy industry. Appl Environ Microbiol. 59(1):197–202. doi:10.1128/aem.59.1.197-202.1993.

Narayanaswamy R, Levy M, Tsechansky M, Stovall GM, O'Connell JD, Mirrielees J, Ellington AD, Marcotte EM. 2009. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. Proc Natl Acad Sci U S A. 106(25):10147–10152. doi:10.1073/pnas.0812771106.

Neugebauerová S, Kypr J. 2000. Invariant and variable base stacking geometries in B-DNA and a-DNA. J Biomol Struct Dyn. 18(1):73–81. doi:10.1080/07391102.2000.10506648.

Noree C, Begovich K, Samilo D, Broyer R, Monfort E, Wilhelm JE. 2019. A quantitative screen for metabolic enzyme structures reveals patterns of assembly across the yeast metabolic network. Mol Biol Cell. 30(21):2721–2736. doi:10.1091/mbc.E19-04-0224.

Noree C, Sato BK, Broyer RM, Wilhelm JE. 2010. Identification of novel filament-forming proteins in saccharomyces cerevisiae and drosophila melanogaster. J Cell Biol. 190(4):541–551. doi:10.1083/jcb.201003001.

Nowotny M, Gaidamakov SA, Crouch RJ, Yang W. 2005. Crystal structures of Rnase H bound to an Rna/DNA hybrid: Substrate specificity and metal-dependent catalysis. Cell. 121(7):1005–1016. doi:10.1016/j.cell.2005.04.024.

Nowotny M, Yang W. 2006. Stepwise analyses of metal ions in Rnase H catalysis from substrate destabilization to product release. Embo J. 25(9):1924–1933. doi:10.1038/sj.emboj.7601076.

Olsen BR, Svenneby G, Kvamme E, Tveit B, Eskeland T. 1970. Formation and ultrastructure of enzymically active polymers of pig renal glutaminase. J Mol Biol. 52(2):239–245. doi:10.1016/0022-2836(70)90028-8.

Olson JA, Anfinsen CB. 1952. The crystallization and characterization of L-glutamic acid dehydrogenase. J Biol Chem. 197(1):67–79. doi:10.1016/S0021-9258(18)55654-4.

Park CK, Horton NC. 2019. Structures, functions, and mechanisms of filament forming enzymes: a renaissance of enzyme filamentation. Biophys Rev. 11(6):927–994. doi:10.1007/s12551-019-00602-6.

Park CK, Horton NC. 2020. Novel insights into filament-forming enzymes. Nat Rev Mol Cell Biol. 21(1):1–2. doi:10.1038/s41580-019-0188-1.

Park CK, Joshi HK, Agrawal A, Ghare MI, Little EJ, Dunten PW, Bitinaite J, Horton NC. 2010. Domain swapping in allosteric modulation of DNA specificity. PLoS Biol. 8(12):e1000554. doi:10.1371/journal.pbio.1000554.

Park CK, Sanchez JL, Barahona C, Basantes LE, Sanchez J, Hernandez C, Horton NC. 2018. The Run-on oligomer filament enzyme mechanism of SgrAI: part 1. Assembly kinetics of the run-on oligomer filament. J Biol Chem. 293(38):14585–14598. doi:10.1074/jbc.RA118.003680.

Park CK, Sanchez JL, Barahona C, Basantes LE, Sanchez J, Hernandez C, Horton NC. 2018. The Run-on oligomer filament enzyme mechanism of SgrAI: part 2. Kinetic modeling of the full DNA cleavage pathway. J Biol Chem. 293(38):14599–14615. doi:10.1074/jbc.RA118.003682.

Park CK, Stiteler AP, Shah S, Ghare MI, Bitinaite J, Horton NC. 2010. Activation of DNA cleavage by oligomerization of DNA-bound SgrAI. Biochemistry. 49(41):8818–8830. doi:10.1021/bi100557v.

Parmeggiani A, Luft JH, Love DS, Krebs EG. 1966. Crystallization and properties of rabbit skeletal muscle phosphofructokinase. J Biol Chem. 241(20):4625–4637.

Pauling L. 1947. Atomic radii and interatomic distances in metals. J Am Chem Soc. 69(3):542–553. doi:10.1021/ja01195a024.

Petrovska I, Nüske E, Munder MC, Kulasegaran G, Malinovska L, Kroschwald S, Richter D, Fahmy K, Gibson K, Verbavatz J-M, et al. 2014. Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. Elife. 3:e02409. doi:10.7554/eLife.02409.

Pingoud A, Fuxreiter M, Pingoud V, Wende W. 2005. Type II restriction endonucleases: structure and function. Cell Mol Life Sci. 62(6):685–707. doi:10.1007/s00018-004-4513-1.

Pingoud A, Jeltsch A. 2001. Structure and function of type II Restriction endonucleases. Nucleic Acids Res. 29(18):3705–3727. doi:10.1093/nar/29.18.3705.

Pingoud A. 2004. Restriction endonucleases. Berlin, New York: Springer-Verlag.

Polley S, Lyumkis D, Horton NC. 2019. Mechanism of filamentation-induced allosteric activation of the SgrAI endonuclease. Structure. 27(10):1497–1507.e3. doi:10.1016/j.str.2019.08.001.

Polley S, Lyumkis D, Horton NC. 2019. *Structure of a DNA-bound dimer extracted from filamentous SgrAI endonuclease in its activated form*.

Roberts RJ, Halford SE, Linn SM, Lloyd RS, Roberts RJ. 1993. Type II restriction endonucleases. In Nucleases. Cold Sping Harbor, NY: Cold Spring Harbor Laboratory Press. p. 35–88.

Sam MD, Horton NC, Nissan TA, Perona JJ. 2001. Catalytic efficiency and sequence selectivity of a restriction endonuclease modulated by a distal manganese ion binding site. J Mol Biol. 306(4):851–861. doi:10.1006/jmbi.2000.4434.

Sam MD, Perona JJ. 1999. Catalytic roles of divalent metal ions in phosphoryl transfer by ecorv endonuclease. Biochemistry. 38(20):6576–6586. doi:10.1021/bi9901580.

Shah S, Sanchez J, Stewart A, Piperakis MM, Cosstick R, Nichols C, Park CK, Ma X, Wysocki V, Bitinaite J, et al. 2015. Probing the Run-on oligomer of activated SgrAI bound to DNA. PLoS One. 10(4):e0124783. doi:10.1371/journal.pone.0124783.

Shan Z, Ghadirian N, Lyumkis D, Horton NC. 2022. Pretransition state and apo structures of the filament-forming enzyme SgrAI Elucidate mechanisms of activation and substrate specificity. J Biol Chem. 298(4):101760. doi:10.1016/j.jbc.2022.101760.

Shan Z, Rivero-Gamez A, Lyumkis D, Horton NC. 2024. Two-metal ion mechanism of DNA cleavage by activated, filamentous SgrAI. J Biol Chem. 300(8):107576. doi:10.1016/j.jbc.2024.107576.

Shen Q-J, Kassim H, Huang Y, Li H, Zhang J, Li G, Wang P-Y, Yan J, Ye F, Liu J-L. 2016. Filamentation of metabolic enzymes in *Saccharomyces Cerevisiae*. J Genet Genomics. 43(6):393–404. doi:10.1016/j.jgg.2016.03.008.

Siksnys V, Skirgaila R, Sasnauskas G, Urbanke C, Cherny D, Grazulis S, Huber R. 1999. The Cfr10i restriction enzyme is functional as a tetramer. J Mol Biol. 291(5):1105–1118. doi:10.1006/jmbi.1999.2977.

Simonet JC, Burrell AL, Kollman JM, Peterson JR. 2020. Freedom of assembly: metabolic enzymes come together. Mol Biol Cell. 31(12):1201–1205. doi:10.1091/mbc.E18-10-0675.

Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A. 90(14):6498–6502. doi:10.1073/pnas.90.14.6498.

Sträter N, Lipscomb WN, Klabunde T, Krebs B. 1996. Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. Angew Chem Int Ed Engl. 35(18):2024–2055. doi:10.1002/anie.199620241.

Suresh HG, da Silveira Dos Santos AX, Kukulski W, Tyedmers J, Riezman H, Bukau B, Mogk A. 2015. Prolonged starvation drives reversible sequestration of lipid biosynthetic enzymes and organelle reorganization in saccharomyces cerevisiae. Mol Biol Cell. 26(9):1601–1615. doi:10.1091/mbc.E14-11-1559.

Szczelkun MD. 2002. Kinetic models of translocation, head-on collision, and DNA cleavage by type I restriction endonucleases. Biochemistry. 41(6):2067–2074. doi:10.1021/bi011824b.

Tautz N, Kaluza K, Frey B, Jarsch M, Schmitz GG, Kessler C. 1990. SgrAI, a novel class-II restriction endonuclease from streptomyces griseus recognizing the octanucleotide sequence 5′-Cr/Ccggyg-3′ [Corrected]. Nucleic Acids Res. 18(10):3087–3087. doi:10.1093/nar/18.10.3087.

Teixeira F, Tse E, Castro H, Makepeace KAT, Meinen BA, Borchers CH, Poole LB, Bardwell JC, Tomás AM, Southworth DR, et al. 2019. Chaperone activation and client binding of a 2-cysteine peroxiredoxin. Nat Commun. 10(1):659. doi:10.1038/s41467-019-08565-8.

Tümmler B. 2014. Genotyping methods. In: Filloux A, Ramos J-L, editors. Pseudomonas methods and protocols. New York, NY: Springer New York. p. 33–47.

Ulyanov NB, Zhurkin VB. 1984. Sequence-dependent anisotropic flexibility of B-DNA. A conformational study. J Biomol Struct Dyn. 2(2):361–385. doi:10.1080/07391102.1984.10507573.

Vagelos PR, Alberts AW, Martin DB. 1962. Activation of acetyl-coa carboxylase and associated alteration of sedimentation characteristics of the enzyme. Biochem Biophys Res Commun. 8(1-2):4–8. doi:10.1016/0006-291x(62)90224-3.

Vanamee ES, Viadiu H, Kucera R, Dorner L, Picone S, Schildkraut I, Aggarwal AK. 2005. A view of consecutive binding events from structures of tetrameric endonuclease Sfii bound to DNA. Embo J. 24(23):4198–4208. doi:10.1038/sj.emboj.7600880.

Viadiu H, Aggarwal AK. 1998. The role of metals in catalysis by the restriction endonuclease bamhi. Nat Struct Biol. 5(10):910–916. doi:10.1038/2352.

Webb BA, Dosey AM, Wittmann T, Kollman JM, Barber DL. 2017. The glycolytic enzyme phosphofructokinase-1 assembles into filaments. J Cell Biol. 216(8):2305–2313. doi:10.1083/jcb.201701084.

Wentzell LM, Nobbs TJ, Halford SE. 1995. The Sfii restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. J Mol Biol. 248(3):581–595. doi:10.1006/jmbi.1995.0244.

Werner JN, Chen EY, Guberman JM, Zippilli AR, Irgon JJ, Gitai Z. 2009. Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. Proc Natl Acad Sci U S A. 106(19):7858–7863. doi:10.1073/pnas.0901781106.

Wilson GG, Murray NE. 1991. Restriction and modification systems. Annu Rev Genet. 25(1):585–627. doi:10.1146/annurev.genet.25.1.585.

Wood KM, Daniels LE, Halford SE. 2005. Long-range communications between DNA sites by the dimeric restriction endonuclease SgrAI. J Mol Biol. 350(2):240–253. doi:10.1016/j.jmb.2005.04.053.

Yang W, Lee JY, Nowotny M. 2006. Making and breaking nucleic acids: two-Mg2+-Ion catalysis and substrate specificity. Mol Cell. 22(1):5–13. doi:10.1016/j.molcel.2006.03.013.

Yang W. 2011. Nucleases: diversity of Structure, Function and Mechanism. Q Rev Biophys. 44(1):1–93. doi:10.1017/S0033583510000181.

Zhong J, Guo CJ, Zhou X, Chang CC, Yin B, Zhang T, Hu HH, Lu GM, Liu JL. 2022. Structural basis of dynamic P5cs filaments. Elife. 11:e76107. doi:10.7554/eLife.76107.