Biochemistry

pubs.acs.org/biochemistry Article

DNA Sequence Control of Enzyme Filamentation and Activation of the SgrAI Endonuclease

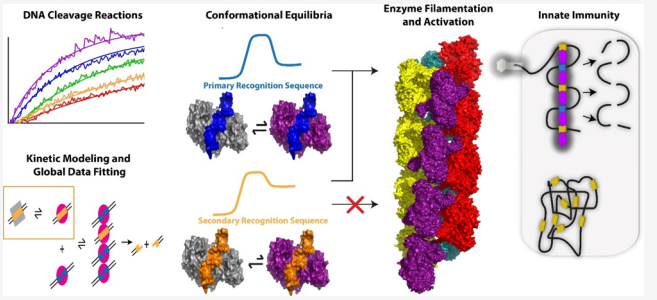
Niloofar Ghadirian, Richard D. Morgan, and Nancy C. Horton*

Cite This: *Biochemistry* 2024, 63, 326−338 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

Downloaded via UNIV OF ARIZONA on February 10, 2024 at 18:11:36 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ABSTRACT: Enzyme polymerization (also known as filamentation) has emerged as a new layer of enzyme regulation. SgrAI is a sequence-dependent DNA endonuclease that forms polymeric filaments with enhanced DNA cleavage activity as well as altered DNA sequence specificity. To better understand this unusual regulatory mechanism, full global kinetic modeling of the reaction pathway, including the enzyme filamentation steps, has been undertaken. Prior work with the primary DNA recognition sequence cleaved by SgrAI has shown how the kinetic rate constants of each reaction step are tuned to maximize activation and DNA cleavage while minimizing the extent of DNA cleavage to the host genome. In the current work, we expand on our prior study by now including DNA cleavage of a secondary recognition sequence, to understand how the sequence of the bound DNA modulates filamentation and activation of SgrAI. The work shows that an allosteric equilibrium between low and high activity states is modulated by the sequence of bound DNA, with primary sequences more prone to activation and filament formation, while SgrAI bound to secondary recognition sequences favor the low (and nonfilamenting) state by up to 40-fold. In addition, the degree of methylation of secondary sequences in the host organism, *Streptomyces griseus*, is now reported for the first time and shows that as predicted, these sequences are left unprotected from the SgrAI endonuclease making sequence specificity critical in this unusual filament-forming enzyme.

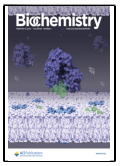


INTRODUCTION

The formation of linear, helical, or cylindrical polymers (i.e., molecular filaments) by metabolic enzymes was discovered over 40 years ago but has not been appreciated as a widespread phenomenon, or as a mechanism of enzyme regulation, until relatively recently. Early studies in the 1970s with various enzymes purified from their natural sources demonstrated oligomerization/polymerization using a variety of biophysical methods.[1−14] Few further studies on this phenomenon were performed until fluorescence microscopy showed unexpectedly that many different cellular enzymes form various superstructures within cells, which are also known as filaments or cytoophidia.[15−20] In some (but certainly not all) cases, it is suspected that the formation of polymeric filaments by particular enzymes stimulates the formation of these large cellular self-assemblies.[21] Altogether, more than 80 distinct proteins are known to form such structures in cells and more than 30 that form polymeric filaments *in vitro*.[21] These enzymes derive from bacteria, plants, yeast, flies, and human[22] and participate in diverse pathways such as carbohydrate, amino acid, fatty acid, and nucleic acid metabolism, as well as in translation, innate immunity, and signaling, among others.[22] Hence, the phenomenon of enzyme self-assembly (molecular and cellular) is widespread, found in diverse cell types, across all domains of life, and in diverse cellular pathways. Why these enzymes form polymeric filaments and/or cellular self-assemblies remains an important question. Some hypotheses include rapid activation or inactivation of enzymatic activity, modulation of allosteric responses, storage and protection from autophagy, control of cellular localization, and/or buffering of enzyme activity in cells.[21,23] While the structures of these filamentous enzymes are now being revealed,[24−33] including for the SgrAI system,[34−36] far fewer studies have been performed to determine the underlying kinetic mechanisms of the phenomenon.

SgrAI is a sequence-specific double-stranded DNA endonuclease from *Streptomyces griseus* (a type II restriction endonuclease, type II REs). In the absence of DNA, SgrAI is a homodimer composed of two identical 37 kDa chains, each with a single active site.[37] The enzyme binds its recognition sequences in double-stranded DNA with nanomolar affinity and in a 1:1 (DNA duplex):(SgrAI dimer) ratio, giving rise to the DNA-bound SgrAI dimer (DBD), which is capable of making endonucleolytic cleavages in each strand of duplex

Received: June 16, 2023
Revised: December 21, 2023
Accepted: December 21, 2023
Published: January 11, 2024



ACS Publications © 2024 American Chemical Society

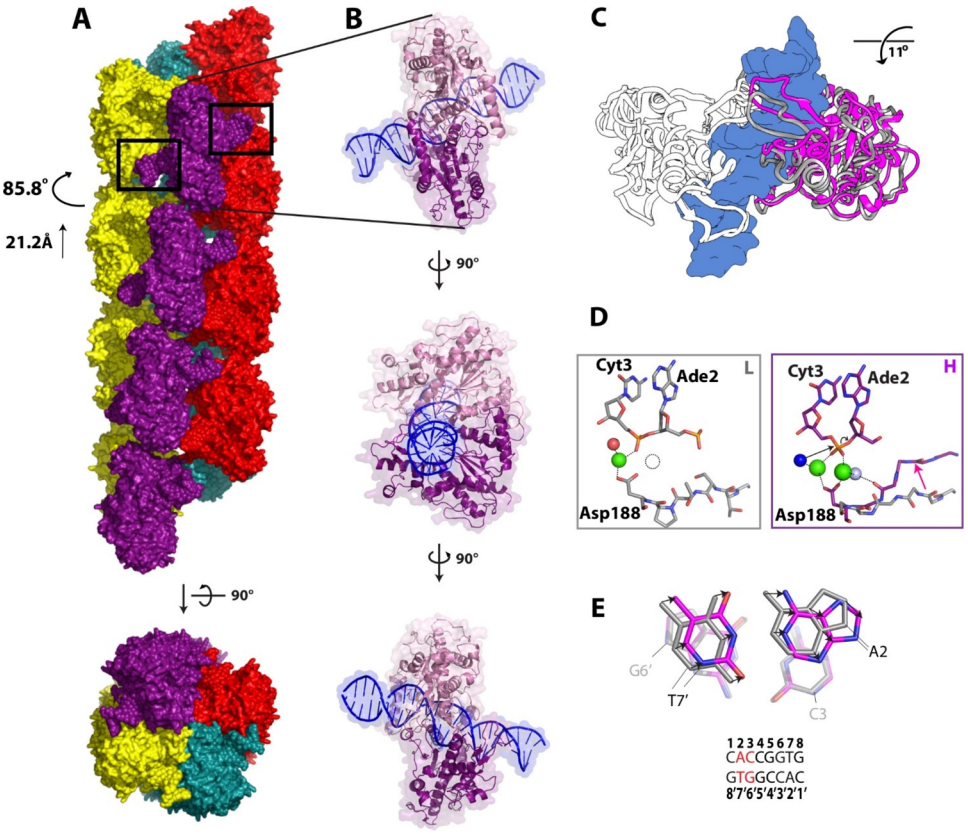


Figure 1. Filamentous form of SgrAI, the DNA bound SgrAI dimer (DBD), and conformational differences between low and high activity states. (A) Filamentous form of SgrAI bound to primary site DNA (PDB entry 7SS5). Each DNA-bound SgrAI dimer (DBD) is colored dark magenta, red, yellow, or teal. Successive DBDs assemble in a left-handed helical fashion with 85.8° rotation and 21.2 Å translation from the DBD before. Black boxed regions show contacts between the flanking DNA of one DBD and two neighboring DBD. (B) A single DBD from the filament is colored to show the individual chains of the SgrAI homodimer (dark magenta, pink) and bound DNA (blue). (C) Comparison of the low activity L state (white/gray, PDB entry 3DVO) and the high activity H state (white/magenta, PDB entry 7SS5) conformations shows an 11° difference in intersubunit rotation. DNA from the H state shown in blue. (D) L (gray, left panel) and H (magenta, right panel) arrangements in DNA cleavage active sites. The subunit rotation of the H state results in creation of a second Mg^{2+} binding site, which likely accelerates the DNA cleavage reaction (arrow shows shift of SgrAI segment). Green, Mg^{2+}, dark blue, nucleophile (water or hydroxide), light blue, water molecule linking the new Mg^{2+} ion to a structurally altered segment of SgrAI (pink arrow). The black arrows indicate the attack on the phosphorus atom by the nucleophile (straight arrow) and bond breakage (curved arrow) steps of the DNA cleavage reaction. (E) Base stacking between the second and third nucleotides of the recognition sequence differs between L (gray) and H (magenta) states. All atoms of the bases in the G6:C3′ base pair were used in the superposition. Arrows shown to emphasize the shift in the base positions.

DNA.[37] SgrAI cleaves two types of recognition sequences: primary sequences, CR|CCGGYG (where R = A or G and Y = C or T, and | denotes cleavage site) and secondary, CC|CCGGYG or **D**R|CCGGYG (where D = A, G, or T).[38,39] Initial observations showed that SgrAI cleaves its primary recognition sequences slowly when present in only a single copy on a DNA molecule but considerably faster when present in two or more copies.[40] In addition, SgrAI was observed to cleave secondary sequences but only when a primary sequence was also present on the same DNA molecule.[38] DNA cleavage reaction measurements using oligonucleotides containing only a single copy of a recognition sequence showed that these effects could be reproduced *in trans* but are dependent on the concentration of SgrAI bound to the primary recognition sequence. Subsequent studies showed that SgrAI bound to primary sequences nucleate filaments (Figure 1A,B), which stabilize an altered conformation of SgrAI with activated DNA cleavage activity.[34,35,37,41] SgrAI bound to the secondary sequences do not form filaments on their own but will join filaments nucleated by SgrAI bound to the primary sequence and thereby become activated.[37] The increase in the DNA cleavage rate of SgrAI in the filament, compared to its nonfilamentous form, has been measured to be 200-fold in the case of the primary sequence and up to 1000-fold in the case of secondary (this larger effect is mainly due to the exceedingly slow cleavage rate of secondary sequences by nonfilamentous SgrAI).[37,39] Interestingly, both uncleaved and cleaved forms of the primary recognition sequence activate SgrAI and induce filamentation to similar degrees, provided that these sequences are embedded within a double-stranded molecule with sufficient base pairs flanking either side of the recognition sequence.[35,37−39] Such flanking sequences are important for stabilizing contacts between SgrAI in the filament (Figure 1A,B).

Structures of SgrAI bound to DNA have been solved in filamentous forms (the high activity or H state, or in prior work, the R state), and in the nonfilamentous, dimeric state (the low activity or L state, or in prior work, the T state).[34,36,42,43] Comparison of the representative L and H states shows an 11° rotation of one chain of the dimer relative to the other along an axis roughly perpendicular to the bound DNA helical axis[42] (Figure 1C). To accommodate this

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

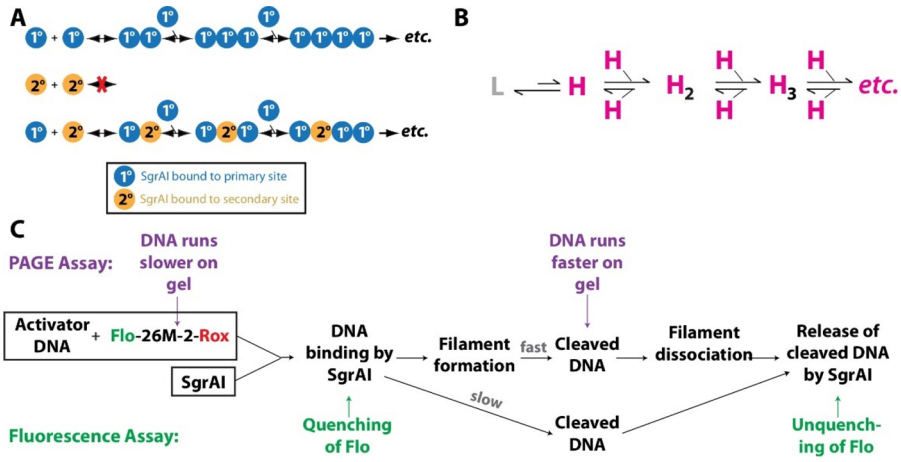


Figure 2. SgrAI mechanism schematics and experimental approach. (A) Schematic showing how SgrAI bound to a primary site sequence (blue) stimulates filamentation, which can incorporate SgrAI bound to a secondary site sequence (yellow); however, SgrAI bound to the secondary site sequence does not stimulate filament formation itself. (B) Allosteric equilibrium and filamentation mechanism. The low DNA cleavage activity L state (gray) is more stable than the high activity H state (magenta). Only the H state forms filaments, thereby stabilizing the H state conformation. (C) Diagram illustrating how DNA cleavage reactions were measured. The reporter (Flo-26M-2-Rox) and activator DNA are mixed, and SgrAI is added to initiate the cleavage reaction. SgrAI rapidly binds both DNAs, and binding to the reporter DNA results in quenching of the Flo emission. Filament formation proceeds, resulting in activated, rapid cleavage of the reporter DNA. Filaments dissociate, allowing the cleaved reporter DNA to dissociate from SgrAI, upon which unquenching of the Flo emission occurs. Independent of filament formation, reporter DNA can also be cleaved by SgrAI although this occurs slowly (~0.0009 s^{-1}).[37] The PAGE assay measures the amount of all cleaved DNA at different times after initiation of the reaction, regardless of whether that cleaved DNA is still bound to SgrAI or within filaments.

intersubunit rotation, residues at the dimeric interface adopt new orientations, which propagate to the active site where DNA cleavage takes place.[34] These shifts create a new binding pocket for a second divalent cation immediately adjacent to the scissile phosphate (i.e., the moiety containing the phosphodiester bond to be cleaved) in each active site (Figure 1D).[34,36] The nonfilamentous structure (the L state, left panel, Figure 1D) shows binding of only a single divalent cation near the DNA; however, two are expected in the two-metal ion mechanism for DNA cleavage.[44−46] Therefore, the second divalent cation that binds only in the filamentous conformation (the H state, right panel, Figure 1D) is expected to be the origin of the accelerated DNA cleavage activity of SgrAI. The conformational change that creates this binding pocket appears to be stabilized via interactions with neighboring DBD in the filament.[34−36] Hence, filamentation stabilizes this activated (H state) conformation of SgrAI, which positions a second divalent cation near the DNA leading to rapid DNA cleavage.[36]

In addition to changes in the structure of SgrAI and the SgrAI-DNA interface, differences in the structure of the bound DNA between the L and H states were also observed.[34] One change that appears especially relevant for the differences in behavior with primary and secondary recognition sequences is a change in base stacking between the second and third nucleotides of the bound recognition sequence (Figure 1E). One of the two types of secondary sequence contains a substitution in the second base pair (i.e., C<u>C</u>CCGGYG) and hence would be impacted by this change in structure.

We also previously conducted a full kinetic analysis to create a computation model for the reaction pathway of primary sequence cleavage by SgrAI.[22,47,48] The model is complete with rate constants for each step of the reaction pathway including filament formation, accelerated DNA cleavage, and filament disassembly. The rate constants derived from global data fitting revealed some unexpected properties of the filament-forming mechanism, which we hypothesize are important for the biological function of SgrAI.[22,47,48] For example, the model showed that the assembly of DNA-bound SgrAI complexes (i.e., DBD) into filaments is slow and can be rate limiting under low concentrations of DBD. However, when DBD concentrations are high, assembly into filaments followed by DNA cleavage (which is 10−40 times faster than filament disassembly) is rapid. The computational model also allowed for modeling conditions within the cell, which show how local concentrations of sites on the same contiguous DNA induce filament formation and DNA cleavage.[22,49] The term "local concentration" refers to the concentration that molecules tethered in close proximity experience (as if they were contained together in a smaller volume determined by the tethering), which is typically much higher than the actual concentration of the molecules.[50] Hence, independent copies of SgrAI bound to different sites on the same DNA molecule are trapped in a smaller volume and are therefore more likely to collide than if bound to separate DNA molecules and are also more likely to form filaments. This effect is important since only primary sites on invading phage DNA will be available for SgAI binding and therefore SgrAI filaments will only form on the invading DNA. Since filamentation also leads to cleavage at the secondary sequences, which are not protected on the host genome (as shown herein for first time), sequestering SgrAI via filamentation on the invading phage is then important for both rapid cleavage of the additional secondary site sequences on the phage DNA and preventing harmful DNA cleavages on the host genome.[22]

Because SgrAI cleaves secondary sequences only in the presence of primary, our model predicts that this is due to the inability of SgrAI bound to secondary sequences to nucleate filamentation (Figure 2A). We reason that binding to the secondary sequences favors the conformation of the enzyme with low DNA cleavage activity, which does not readily form filaments (i.e., the L state, Figure 2B). In order to test this hypothesis, a full kinetic analysis of the cleavage of secondary

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

site sequences by SgrAI was performed and described herein. As before, in studies with primary sequence (a.k.a. primary site) DNA,[47,48] two approaches to measure DNA cleavage in the presence of varying concentrations of activator DNA (a precleaved 40 bp primary site, which induces filamentation) were used. The data were then fit to a computational model developed previously for SgrAI cleavage of the primary sequence,[47,48] using the same rate constants, but with one important difference. To account for the slower cleavage of secondary sequences by SgrAI, and consistent with our hypothesis that the base pair substitutions in the secondary sequences alter the energy landscape of the L and H conformations of the SgrAI/DNA complexes, a new mechanistic step was included that corresponds to the equilibrium between the L and H conformations in the SgrAI/DNA complex prior to filamentation. Fitting of the new experimental data described herein to the modified mechanistic model, combined with other considerations, showed that the secondary sequence shifts the L/H equilibrium toward the L state by a factor of up to 40-fold, compared to that for primary site-bound SgrAI. We discuss the significance of this value with respect to possible structural origins for this preferential L state stability in SgrAI/DNA complexes containing the secondary sequence.

MATERIALS AND EXPERIMENTAL DETAILS

Methylation Analysis of *S. griseus* Genomic DNA. Genomic DNA was isolated from *S. griseus* NEB1061 expressing the SgrAI RM system and used to make a standard PacBio SMRTbell library.[51] The library was sequenced on a PacBio RSII instrument using P6/C4 chemistry yielding a 1068 Mb sequence and assembled into a complete chromosome of 7,960,167 bp. Methylated motifs were identified as previously described using RSII_Modification_and_motif_analysis.1 within SMRT Analysis 2.3.0.[51] Two motifs having m6A methylation were identified with nearly 100% of the sites identified as modified, while the program identified an unknown type of modification at just 29% of a recognition motif that is a subset of the SgrAI motif of CRCCGGYG. This result is typical for 5mC detection by PacBio sequencing, where the 5mC methylation causes a slight perturbation of the polymerase that results in a diffuse and small kinetic signal that is difficult for the software to identify.

To better identify the location of 5mC methylation in the SgrA genome, the PacBio library was treated with TET2 enzyme (NEB E7130) to oxidize 5mC (to 5hmC to 5fC) to 5caC to enhance the kinetic signal in PacBio sequencing.[52] The TET2-treated library was then sequenced on the PacBio RSII to generate 858 Mb of sequence, and modified motifs were identified using the same software. The kinetic signature for 5caC bases is typically largest at the base position two bases 5′ to 5caC (i.e., the original 5mC modified base), with some signal at the modified base. The correct CRCCGGYG (primary recognition sequence) was clearly identified by the software following TET oxidization, with the largest kinetic signal on the C at position 1, indicating that native SgrAI methylation modifies the C at position 3 in the recognition motif: CR**C**m**C**GGYG. The data were further analyzed for the kinetic signals at the two types of secondary sequence as well as a noncognate sequence serving as the negative control.

Protein Purification. SgrAI (UniProt ID Q9F6L0_STRGR) was expressed with a C-terminal his tag in Tuner (DE3) *Escherichia coli* , which also contained the pLysS plasmid (Novagen, Inc.) and the MspI.M expression plasmid (pBAK.MspI).[42] The protein was purified using Talon metal affinity resin (Clontech, Inc.) followed by ion-exchange FPLC using ion-exchange (HiPrep Heparin FF 16/10 column, Cytiva) and size exclusion chromatography (Superdex 200 10/300 GL, Cytiva). Purity of the protein was confirmed using SDS-PAGE. The purified protein was then aliquoted into single-use aliquots, flash frozen in liquid nitrogen, and stored at −80 °C.

DNA Preparation. The oligonucleotides, some of which possess a covalently linked 6-carboxyfluorescein (Flo) connected to the 5′ phosphate via a trans-4-amino cyclohexanol linker or 5(6)-carboxy-X-rhodamine connected to the 5′phosphate via a 6-amino hexan-1-ol linker (Rox), were prepared synthetically and purified (via HPLC or PAGE) from a commercial source (Sigma-Genosys, Inc., or IDTDNA, Inc.). The concentrations were measured spectrophotometrically, with an extinction coefficient calculated from standard values for the nucleotides and any relevant fluorophores.[53] The self-complementary DNA strands, or equimolar quantities of complementary DNA, were annealed by heating to 90 °C for 10 min at a concentration of 0.1−1 mM, followed by slow cooling to 4 °C over 4−5 h in a thermocycler or heat block. Sequences of the DNA used are shown below (recognition sequences are shown in red, green font with yellow highlight indicates secondary sequence substitution):

Reporter DNA. Flo-26M-2-Top: Flo-5′-AATATATAACACCGGGGATAATATTA-3′

Rox-26M-2-Bot 3′-TTATATATTGTGGCCCCTATTATAAT-5′-Rox

Activator PC DNA. PC-top 5′-GATGCGTGGGTCTTCACA −3′

PC-bot 3′-CTACGCACCCAGAAGTGTGGCC-5′

Control and PAGE Marker DNA. Flo-26M-2-Top-Cut: Flo-5′-AATATATAACA-3′

26M-2-Bot-Cut 3′-TTATATATTGTGGCC-5′

26M-2-Top-Cut: 5′-CCGGGGATAATATTA-3′

Rox-26M-2-Bot-Cut: 3′-CCTATTATAAT-5′-Rox

Equilibrium Dissociation Constant Measurements Using Fluorescence Anisotropy. Titrations to determine the equilibrium dissociation constant, K_D, of DNA strand annealing as well as binding of SgrAI to the cleaved DNA oligonucleotides were performed using a fluorescence polarization anisotropy (FPA) technique.[37] Flo-labeled DNA (50 nM) was placed in 1.5 mL of reaction buffer without $MgCl_2$ (50 mM Tris−HCl, 150 mM NaCl, 10% glycerol, 1 mM DTT, pH 8.0 at room temperature), as noted, in a 2 mL cuvette. The cuvette was kept at 25 °C using water circulation in the cuvette holder from a temperature-controlled water bath. An ISS PC1 fluorimeter with polarizers and 0.5 mm slits was used to measure the anisotropy of fluorescence emission at 520 nm with excitation at 495 nm and constant stirring of the sample. The anisotropy of the fluorescence signal was averaged over 100 s in 0.1 s measurements. The data were fit in GraphPad Prism (GraphPad Software, Inc.) using the following equation:

$$A = A_{min} + (A_{max} - A_{min})\left(\frac{X}{(K_D + X)}\right)$$

where A is the anisotropy of a solution with the concentration X of the titrant, A_{min} and A_{max} are the baseline and maximum anisotropy with 0 and 100% binding, respectively, of the species being titrated (i.e., the 50 nM DNA), and K_D is the equilibrium dissociation constant fit in the analysis.

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

Single Turnover DNA Cleavage Assays Measured with Fluorescence. Reactions were carried out in 1.5 mL of reaction buffer with $MgCl_2$ (50 mM Tris−HCl, 150 mM NaCl, 10% glycerol, 10 mM $MgCl_2$, 1 mM DTT, pH 8.0 at room temperature) in a 2 mL cuvette with constant stirring at 25 °C. Each reaction contained 50−100 nM Flo-26M-2-Rox and 0−1 μM unlabeled PC DNA and was initiated by the addition of 0.50−2.5 μM SgrAI. Fluorescence was measured with an ISS PC1 fluorimeter with 495 nm excitation and emission monitored at 520 nm emission via a monochromator and 1 mm (8 nm spectral width) slits. Intensity measurements were taken in 0.1 ms readings and averaged over 10 iterations for readings approximately every 1.1 s.

Single Turnover DNA Cleavage Measured with Denaturing PAGE. Reactions were performed identically to those described above; however, to measure the total amount of cleaved DNA, aliquots of the reaction were taken at different times after initiation (10 s to 50 min) and mixed 1:1 with a quench solution containing 80% formamide and 50 mM EDTA. Aliquots were analyzed by electrophoresing on 20% acrylamide (19:1 acrylamide:bis-acrylamide) gels containing 4 M urea and 1× TBE (89 mM Tris base, 89 mM boric acid, 2 mM EDTA) and scanned with a Pharos phosphoimager (Bio-Rad, Inc.) for Flo and Rox fluorescent bands. Bands were integrated for their intensity using ImageLab (BioRad, Inc.). The percentage of DNA cleaved was calculated by dividing the intensity of the cleaved DNA band by the sum of the intensities of the cleaved and uncleaved DNA bands. This data, as a function of time after initiation, was used in subsequent analyses and data fitting.

Kinetic Modeling and Global Data Fitting. Global data fitting was performed with Kintek Global Kinetic Explorer version 6.3.180116 (Kintek Global Kinetic Explorer Corp.).[54−56] The equations used in the model corresponding to the reaction steps are provided in the Supporting Information (Table S2). Each reaction step has a forward and a reverse rate constant, which may be held fixed or allowed to be fit by the software. Kintek Global Kinetic Explorer (Kintek GKE) calculates the concentrations of each defined species at a given time after initiation of the reaction by numerically solving the differential equations for the forward and reverse reactions of each mechanistic step, and estimated concentrations of species initially derived from the starting concentrations of reactants in each experiment.[54−56] The experimental data are loaded into the software, and the software adjusts the rate constants to find the best fit of the simulated signal to the experimental data.

Normalization of the experimental data was performed to allow equal weighting of all data sets during fitting. As described in the Supporting Information, the change in the fluorescein emission was found to be due to binding of the reporter DNA by SgrAI. After cleavage by SgrAI, the cleaved DNA is released and dissociates into cleaved half-sites, which then further dissociates into single-stranded DNA. The fluorescence from uncleaved, cleaved, and dissociated half-site or single-stranded Flo-labeled DNA was found to be essentially equivalent (see the Supporting Information, Figures S1 and S2). Incorporation of reporter-bound SgrAI into filaments also does not alter the Flo emission (see the Supporting Information, Figure S3A). Hence, the initial drop in Flo emission in reactions is due to binding of the DNA by SgrAI, and the slow recovery is due to the release of cleaved DNA from SgrAI, which remains dissociated as cleaved half-sites and/or its composite single strands. Some rebinding of the cleaved DNA is likely to occur, although inclusion of this process in the model had no effect on the final fitted rate constants (see the Supporting Information and Figures S4 and S5).

The simulated signal for reaction progression in the fluorescence data sets was calculated using the simulated concentrations of species from the model and fitted baseline and scale factors:

$$\text{signal} = \text{baseline} + (\text{scaling factor}) \times [\text{free cleaved DNA}] \quad (1)$$

where the "free cleaved DNA" is all versions of the cleaved reporter DNA not bound to SgrAI, including the annealed cleaved Flo-26M-2-Rox, the two dissociated but duplexed half-sites, and their dissociated single strands.

In the case of the PAGE data, where the percentage of cleaved DNA in the experimental data set is simulated by the Kintek model, the simulated signal is calculated using

$$\text{signal} = \text{baseline} + (\text{scaling factor}) \times [\text{cleaved reporter DNA}]/[\text{total reporter DNA}] \quad (2)$$

where "cleaved reporter DNA" and "total reporter DNA" include all cleaved and total reporter DNA, respectively, whether bound to SgrAI or not because SgrAI are denatured prior to measuring the cleaved and uncleaved DNA in the PAGE analysis.

Equations used for global data fitting are those used in the prior study in the 4EO model,[48] with two exceptions. The first is the inclusion of the conformational equilibrium step of complexes of SgrAI bound to secondary sequences between low (L state) and high (H state) activity conformations. The second is the inclusion of a "backbinding" step allowing for rebinding of the cleaved reporter DNA. As described in the Supporting Information, this step had no effect on the final derived rate constants (values tested range from 1×10^{-6} to 1 nM^{-1} s^{-1}). The rate constants derived in the prior study (investigating the activated cleavage of the primary sequence in a reporter DNA)[48] for all reactions steps, shown in Table 1, were kept constant to allow for the estimation of the equilibrium constant between the L and H states of SgrAI when bound to the secondary recognition sequence.

The 4EO model[48] consists of the following features. Due to the complexity of the formation of a potentially infinitely long filament and the limitations of the approach to modeling the reaction, filaments were necessarily limited to the size of four SgrAI/DNA complexes (i.e., DBD). Previous studies showed that this limitation provided the same estimates of rate constants for each modeled step as a model with longer filaments (up to five DBDs per filament), provided that the concentrations of activator DNA were ≤1 μM (as is the case here).[47,48] Upon addition of SgrAI to the reaction mixture containing DNA, SgrAI binds to DNA in a 1:1 (SgrAI dimer)/(DNA duplex) in a fast, reversible reaction. This binding reaction is too fast to allow fitting of its forward and reverse rate constants by the current experimental data. Therefore, its forward rate constant is set to the diffusion limit of 1 nM^{-1} s^{-1} and its reverse rate constant calculated by the relevant equilibrium dissociation constant measured previously.[37] The activator DNA, PC, is a precleaved 40 bp DNA, which associates via its "sticky" ends (i.e., single-stranded 5′-CCGG-

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

Table 1. Fixed and Fitted Rate Constants[a]

reaction step	reaction	forward rate constant	reverse rate constant
1	annealing of activator DNA	0.02 nM^{-1} s^{-1}	8 s^{-1}
2	binding of SgrAI to activator DNA to create DBD^A	1 nM^{-1} s^{-1}	0.06 s^{-1}
3	binding of SgrAI to reporter DNA to create DBD^S	1 nM^{-1} s^{-1}	0.26 s^{-1}
4	L/H equilibrium with DBD^S	**1.0 s^{-1}**	**5.9 s^{-1}** (5.4–6.6 s^{-1})[b]
5	association of DBD^A to DBD^A	3.0 × 10^{-4} nM^{-1} s^{-1}	0.017 s^{-1}
6	association of DBD^S to DBD^A	2.0 × 10^{-4} nM^{-1} s^{-1}	0.08 s^{-1}
7	cleavage of DNA by DBD^S in a filament	0.8 s^{-1}	0 s^{-1}[c]
8	release of cleaved reporter DNA from nonfilamented DBD^S	>0.4 s^{-1}	ND[d]
9	cleavage of reporter DNA by nonfilamentous DBD^S	0.0009 s^{-1}	0 s^{-1}

[a]Values in bold indicate those fit in the current work. All other values derive from prior work. [b]0.95 Chi^2 confidence boundary. [c]The step was assumed to be irreversible and therefore the reverse rate constant set to 0 s^{-1}. [d]Not determined: values between 1 × 10^{-6} and 1.0 nM^{-1} s^{-1} did not impact fitted rate constants.

3′ overhangs) and binds to SgrAI in a manner equivalent to an uncleaved version to give DBD^A, and these steps are included in the kinetic model, as described previously.[47,48] SgrAI also binds to the reporter DNA (Flo-26M-2-Rox) to produce DBD^S. Filaments are produced by the association of DBD^A with each other and/or with DBD^S. In addition, each DBD (DBD^A and/or DBD^S) has two interfaces, which each can associate with another DBD; however, DBD^S cannot associate directly to another DBD^S. The origin of this rule is in the shorter DNA sequences on either side of the recognition sequence of the reporter DNA; longer sequences such as those in the activator DNA are necessary to stabilize contacts between DBD in the filament (Figure 1A).[35] Because each DBD has two interfaces to bind to other DBD, and filaments build by the sequential assembly of DBD to either end of the filament and with few interactions to DBD beyond those immediately before and after it in the filament, filaments can grow via the addition of singular DBD at either end and without cooperativity in binding (i.e., a DBD has the same affinity for a filament of size 2 as it does for a filament of size 3).

Within the filament, the high activity (H state) conformation of SgrAI is stabilized,[35] which cleaves DNA rapidly in an irreversible step. Dissociation of DBD, including those DBD^S containing cleaved reporter DNA, occurs in a reversible reaction with the same rate constants as association of DBD^S containing uncleaved DNA. Cleaved DNA can be dissociated from SgrAI once the DBD is released from the filament, and the cleaved DNA further dissociates into the two cleaved half-sites and then into single strands. This latter step is greatly favored and in prior studies was modeled as an irreversible step. However, in this work, due to the longer DNA oligonucleotide used (26 vs 18 bp in the prior study), some reassociation and rebinding of this cleaved DNA to SgrAI was found to occur (see the Supporting Information). Further analysis indicated that such "backbinding" does not affect the globally fit rate constants (only the scaling factors for the free cleaved DNA in the fitting of the fluorescence signals).

The new reaction step added in the current work to the 4EO model previously developed,[48] and described above, is the equilibrium between the low activity (L state) and high activity (H state) conformations of DBD^S. When SgrAI initially binds the reporter DNA (Flo-26M-2-Rox) containing the secondary sequence, it is the low activity L state. This state will not cleave DNA and will not assemble into filaments. It is also favored over the H state, yet when the DBD^S does assume the H state, it may join filaments where this H state is stabilized via interactions with the other DBD of the filament. It was assumed that the interconversion between conformations is rapid; hence, a value of 1 s^{-1} was set for the conversion of the L to the H state. The reverse rate constant was allowed to vary during global data fitting allowing the ratio of the two rate constants to be determined.

Error Analysis. The Fitspace module of Kintek Global Kinetic Explorer version 6.3.180116 was used to determine boundaries (i.e., error limits) for fitted rate constants at the recommended "0.95 χ^2 threshold".[54] Fitspace varies the rate constants systematically while simultaneously fitting all other fitable parameters and recalculating χ^2, a measure of how well the simulated curves match the experimental data (i.e., the sum of the squares of the residuals, with a smaller number indicating a better fit). Values for the rate constants giving a χ^2 withing 5% of the minimum χ^2 were considered within the 0.95 χ^2 threshold.

Calculation of R^2 between Simulated Signal and Experimental Data. Microsoft Excel (Microsoft, Inc.) was used to calculate R^2 between the simulated signal from Kintek modeling and the experimental data. Linear interpolation using the INDEX and MATCH commands was used to determine the dependent variable value (i.e., the simulation data value) to match the exact dependent variable value (i.e., time in seconds) of the experimental data set, and then the RSQ function was used to calculate the R^2.

RESULTS

Single Turnover DNA Cleavage Reactions. DNA cleavage reactions were measured using two types of reactions (summarized in Figure 2C). The first assay measures the changes in the fluorescence signal of a fluorophore-labeled reporter DNA (containing the secondary sequence) as a function of time after the initiation of the reaction (by the addition of SgrAI) (green, Figure 2C). The second assay measures the percentage of reporter DNA cleaved at various times after reaction initiation using denaturing gel electrophoresis (purple, Figure 2C). These two assays provide unique information on the reaction, because they report on a distinct set of steps of the reaction pathway (Figure 2C). The reporter DNA consists of a 26-base-pair duplex DNA (Flo-26M-2-Rox, see Section 2) containing a secondary sequence of the A type, namely, C<u>C</u>CCGGTG (the secondary sequence substitution is underlined), which is 5′ labeled on the top strand with fluorescein (i.e., Flo) and 5′ labeled on the bottom strand with rhodamine-X (i.e., Rox). Although initially designed to detect DNA cleavage via the loss of FRET between the fluorescein and rhodamine-X fluorophores, it was instead discovered through a series of control experiments that changes in fluorescein emission over the course of the reaction were due to SgrAI binding and unbinding and not to changes in FRET

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

between the Flo and Rox fluorophores (see the Supporting Information and Figures S1−S5). The control experiments also suggest some "backbinding" of the cleaved reporter DNA to SgrAI but also showed that the rate constants derived from global fitting of the experimental data to the Kintek model are not impacted by this process. Details of the data sets measured using fluorescence emission are summarized in Table S1 , and the data are plotted in Figure S6.

The second approach to measuring DNA cleavage rates utilized the same reaction compositions as described above, but rather than measuring changes in Flo emission, the percentage of cleaved DNA was measured via denaturing PAGE. The cleaved top and bottom strands migrate differently in such gels, and their different fluorophore labels (fluorescein and rhodamine-X) were individually quantitated (Figure S7). No significant differences in cleavage rates of the two strands were observed (Figure S8), despite their inherent asymmetry in sequence (the secondary sequence substitution is adjacent to the cleavage site on the bottom strand, but five nucleotides away on the top strand; see Section 2 for DNA sequences).

A total of 17 different DNA cleavage reactions were measured via the two types of assays and used in global data fitting to the kinetic model for DNA cleavage by SgrAI. All reactions contained an excess concentration of SgrAI (typically 1 μM) to ensure complete binding of both the reporter and activator DNA. The reporter DNA concentration was kept low (typically 50 nM), and varied concentrations of the activator DNA (50−500 nM) were used in each experiment to provide different degrees of activation of SgrAI (Figure 3A−C and Table S1). The activator DNA is a precleaved 40 bp DNA duplex containing a single primary site sequence (PC DNA)(see Section 2). This DNA anneals and binds SgrAI to nucleate filaments, which then recruit SgrAI bound to the reporter DNA containing the secondary sequence. Once in the filament, SgrAI is activated and rapidly cleaves the secondary sequence. Increasing concentrations of activator DNA lead to increasing degrees of filament formation and faster DNA cleavage (Figure 3A−C).

In addition to modeling the experimental data using the program Kintek Global Kinetic Explorer (Kintek GKE), each data set was also analyzed independently by fitting to a single exponential function (Table S1). The resulting single exponential rate constants are plotted vs the concentration of activator PC DNA in Figure S9. A trend of increasing rate constant is apparent with increasing activator PC DNA concentrations. The data were divided into those from the PAGE measurements (red circles, Figure S9) and those from fluorescence dequenching (blue triangles, Figure S9). Attempting to fit a line to each type of data in Figure S9 shows that the data from PAGE are more scattered, consistent with the noisier data in the PAGE data itself (Figure 3B,C and Figure S8). The rate constants are also higher when measured using the PAGE assay than the fluorescence assay at PC DNA concentrations at 150 nM or above (compare red circles to blue triangles, Figure S9). At the lowest concentrations of PC DNA (50−100 nM), the rate constants calculated using the two approaches are very similar. This can be explained by the fact that at lower concentrations of PC DNA, filament assembly is rate limiting, which affects both experimental approaches to measure DNA cleavage (see Figure 2C). At higher concentrations of PC DNA, later steps become more rate limiting, such as dissociation of the SgrAI DBD from the filament. Data collected via the PAGE method does not include this step and hence shows a faster rate. Therefore, the data from both types of measurements are important for constraining rate constants in the kinetic model. Both assays constrain the rate constants for the filament association step, particularly at low concentrations of activator PC DNA. The PAGE data provides limits on the rate constant for DNA cleavage in the filament, particularly at higher concentrations of PC DNA. The fluorescence data constrains rate constants for filament disassembly, which is a necessary step before the cleaved DNA can be released.[48]

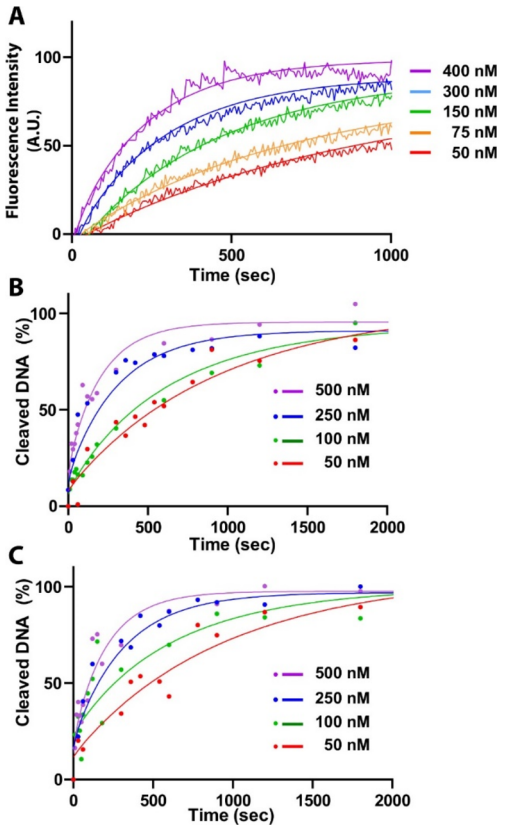


Figure 3. Experimental data and simulated reaction progress curves from global modeling. (A) Representative fluorescence (jagged lines) and simulated (smooth curves) data colored coded by activator PC DNA concentration as shown. The *y*-axis shows the normalized fluorescein (Flo) fluorescence of the reporter DNA (Flo-26M-2-Rox) as a function of time after the reaction initiation. (B) Representative DNA cleavage data measured using gel electrophoresis (filled circles) and simulated data (lines) for all Rox-labeled strands. Data are shown as filled circles, and simulations are shown as smooth curves. Color coded by activator PC DNA concentration, as indicated. (C) As in (B), but Flo-labeled strand. The similarity of the simulated data from the global fit to the time-course data in panels (A−C) shows the robustness of the model.

Kinetic Modeling and Global Data Fitting. Figure 4 shows a schematic of a greatly simplified version of the full reaction pathway of SgrAI used in the computational model (see Table S2 and Section 2 for equations used in modeling, and Table 1 for final rate constants).[56] To construct the reaction model, each step of the reaction pathway is defined with the corresponding forward and reverse rate constants. The first step in the reaction model is the annealing of the two halves of the activator DNA into one 40 bp duplex with a single, complete primary sequence but that contains a missing phosphate moiety at the cleavage site (blue, Figure 4).

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

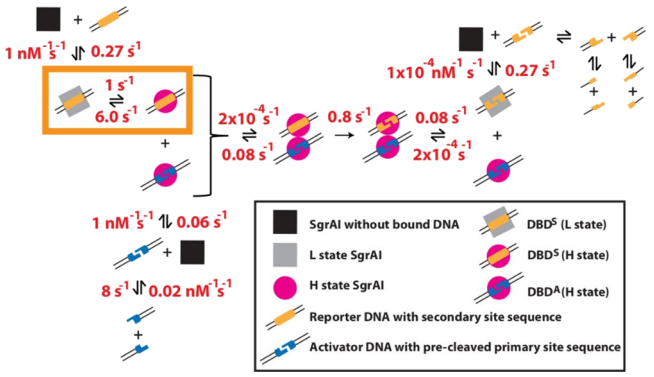


Figure 4. Simplified mechanistic model of that used in global fitting. SgrAI (black square) binds to either reporter DNA containing the secondary sequence (yellow) or activator DNA containing a precleaved version of the primary sequence (blue). The DNA-bound SgrAI is in equilibrium between the low-activity L state (gray box) and the high-activity H state conformation (magenta circle). Only when it is in the H state can it associate with SgrAI bound to activator DNA (alone or in a filament). Although this equilibrium is believed to exist for SgrAI bound to either type of DNA sequence, it is explicitly modeled only for the complex containing the secondary sequence (yellow boxed equilibrium). For simplicity, only one type of filament is shown, with one copy each of SgrAI bound to secondary or activator DNA, although filaments containing up to four copies of SgrAI bound to DNA were used in modeling. Within the filament, DNA cleavage is accelerated and the secondary site DNA is cleaved. Filaments dissociate in a reversible step, followed by dissociation of the cleaved secondary site DNA. The cleaved DNA dissociates into the two half-sites, which further dissociate into single strands Not shown but included in the Kintek model is the slow DNA cleavage by nonfilamented SgrAI.

Previous studies have shown that the absence of this phosphate has minimal effect on the binding affinity of SgrAI for the DNA.[37] This duplex binds SgrAI (black square, Figure 4) to form DBD^A, which is a DNA-bound SgrAI dimer containing the activator DNA. DBD^A assemble into filaments, which in principle are not limited in length, but filaments up to four DBD only were modeled due to computational limits of the modeling approach. Prior work has demonstrated that models with filament lengths of four DBD simulate experimental data to the same degree of accuracy as those that include five DBD and that reactions with ≤ 1 μM of activator DNA result in filaments mostly composed of four DBD or smaller.[47,48] For simplicity, the model in Figure 4 shows only a filament of size 2. The reporter DNA, containing the fluorescein and rhodamine-X fluorophores (yellow, Figure 4), binds SgrAI to create DBD^S. The new step introduced in the current work is boxed in yellow in Figure 4, and it consists of an equilibrium between two conformational states (L and H) of DBD^S. In this diagram, the L state of SgrAI bound to DNA is represented by a gray box, the H state is represented by a magenta circle (Figure 4), and only the H state can be incorporated into filaments. However, DBD^S will not bind each other to form filaments even in the H state because the length of DNA flanking the recognition sequence in the reporter DNA is too short to stabilize contacts between DBD in the filament (see Figure 1A).[35,37] DBD^S can bind to DBD^A to create a "filament" of size 2, or they can join longer filaments by attaching to either of its ends, but only if DBD^A (and not DBD^S) is present at that filament end. DBD^S in the H state cleaves the bound reporter DNA rapidly. Filaments dissociate, and cleaved DNA is released from SgrAI.

All steps of the reaction are modeled as reversible with the exception of the DNA cleavage step and hence are governed by both forward and reverse rate constants. It is these rate constants that are adjusted by the modeling software to simulate reaction curves that best match the experimental data. However, because the goal in this work was to test if the model (and rate constants) determined using experimental data for the cleavage of the primary sequence could be used to accurately predict the cleavage data for a secondary sequence, the same rate constants used in that earlier work[48] were used here. Cleavage of the secondary sequence by SgrAI is slower than cleavage of the primary sequence and requires a greater concentration of activator DNA in these assays (as observed previously[37,39]) to achieve the same rates of DNA cleavage as the primary sequence. We previously hypothesized that the slower cleavage of secondary sequences, and its inability to serve as activating DNA for SgrAI filamentation,[39] was due to the greater stabilization of the nonfilamentation-competent L state of SgrAI. Hence, in this work, an additional step (yellow boxed region in Figure 4) was included describing an equilibrium between L and H states of nonfilamentous DBD^S. Because the *relative* stabilities of these two states is of interest, the forward rate constant (i.e., for the L to H transition) was held constant at 1 s^{-1} and the reverse was varied to find the lowest χ^2 (a measure of the agreement between model simulations and experimental data) in fitting all 17 independent data sets simultaneously. In addition to varying this rate constant, scaling factors that adjust the simulated concentrations of product species (total cleaved DNA in the case of the PAGE data sets or released cleaved DNA in the case of the fluorescence data sets) to the experimental data were also allowed to vary (shown in Table S1). As a result, the value of the H to L rate constant giving the lowest χ^2 for the comparison of the simulated signal and the experimental data was found to be 5.9 s^{-1}, with values from 5.4 to 6.6 s^{-1} giving χ^2 within 5% of the best χ^2. In addition, the R^2, describing the goodness of fit of the simulated curves to the experimental data, were excellent for most fits with values greater than 0.95 (an average of 0.94 with standard deviation of 0.05; Table S1 and Figures S6 and S8).

The Length of the Reporter DNA Matters. In this study, the rate of cleavage by SgrAI of a secondary sequence embedded in a 26 bp reporter DNA was measured and compared to that of a primary sequence; however, the primary sequence used in the prior study was embedded in a much shorter reporter DNA of only 18 bp.[48] The shorter reporter DNA was necessary in a prior study in order to control the degree of SgrAI activation (and filamentation). The 18 bp reporter DNA contains the 8 bp primary recognition sequence flanked by 5 bp on either side, which are too short to stabilize DBD interactions in filaments (see boxed regions in Figure 1A) and thereby allowing the basal rate of DNA cleavage by SgrAI in the nonfilamented form to be measured. In addition, the degree of activation (and filamentation) can be controlled by varying the concentration of added PC DNA. PC DNA is a primary sequence embedded in 40 bp DNA with 16 bp flanking either side of the recognition sequence, long enough to stabilize DBD interactions in filaments.

A systematic study showed that at least 9 flanking bp are necessary to provide robust activation of primary sequence cleavage by SgrAI.[35] The bp flanking a recognition sequence

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

contacts neighboring DBD in the filament (Figure 1A,B) and likely increases the affinity between DBD in the filament. The degree to which the additional flanking bp increases the affinity of DBD for the filament can be estimated; prior studies using an approach to the equilibrium method and kinetic modeling was used to derive the forward and reverse rate constants for SgrAI bound to primary sequence DNA with either 5 or 16 bp flanking DNA (see also the ratio of reverse to forward rate constants for steps 5 and 6 in Table 1).[47] These rate constants predict a K_D of 56 nM for two DBD each with a primary sequence flanked by 16 bp and 400 nM for the interaction between one DBD with 5 flanking bp and one with 16. Hence, the additional 11 bp of flanking DNA results in a sevenfold difference in affinity between two DBD in a filament.

The current study utilized a secondary sequence embedded in a reporter DNA 26 bp in length and therefore contains 9 flanking bp. It was necessary to incorporate the four additional flanking bp because a secondary sequence in an 18 bp reporter DNA could not be activated sufficiently even at the highest concentrations of PC DNA.[37] Therefore, to determine the true effect of the secondary sequence substitution on the relative L and H stabilities, both effects must be combined (i.e., the sixfold effect from kinetic modeling and the up to sevenfold effect from the adding flanking bp on the reporter DNA). Together, these indicate up to a 40-fold effect.

Methylation of SgrAI Recognition Sequences in *S. griseus*. SgrAI is a type II restriction endonuclease and as such is coexpressed with a cognate (i.e., acting on the same recognition sequence) methyltransferase, M.SgrAI. Such a methyltransferase should methylate a base of the endonuclease recognition sequence, rendering it uncleavable by the cognate endonuclease. The M.SgrAI enzyme has been shown to methylate the third nucleotide of the recognition sequence (i.e., CR**C**^{m}CGGYG) at its 5′ position, and by inference, this modification is expected to inhibit cleavage by the SgrAI endonuclease.[57] However, since the endonuclease SgrAI was found to also cleave the additional secondary sequences (C**C**CCGGYG and **D**RCCGGYG), it was of interest to learn if these sequences might also be methylated similarly and thereby protected from cleavage by SgrAI. Hence, a DNA sequencing technique to determine the methylation status of secondary sequences, as well as primary (as a positive control) and unrelated sequences (as a negative control) in the organism hosting the SgrAI system (with endonuclease and methyltransferase), *S. griseus*.

PacBio sequencing of TET2-treated genomic DNA was performed (see Section 2), which detects methylation of the C5 position of cytosine by slower base incorporation (IPD) at the methylated C as well as the base position two bases 5′ of the methylated C. Methylation is expected at the second C, in the third position of the recognition sequence (CR**C**^{m}CGGYG); hence, the signal is expected at this position as well as that 2 nt 5′, namely, the first C of (C**R**C^{m}CGGYG). Strong signals were found at these two positions in primary recognition sequences (black, Figure 5), with 94% of reads having an IPD greater 4.8. Incomplete TET conversion could account for the remaining 6%, thus indicating 100% methylation of primary sites. In contrast, little to no methylation is detected in secondary (blue and red, Figure 5) and the average IPD values of the two types of secondary sites are 1.0 and 1.3 for types A (C**C**CCGGYG) and B (**D**RCCGGYG), respectively. In the case of a noncognate sequence (**D**RCCGGY**H**, H = A, C, or T), serving as a negative control (green, Figure 5), IPD is also 1.0, indicating the absence of detected methylation. Thus, the SgrAI methyltransferase appears highly specific for the primary recognition sequences CRCCGGYG, leaving secondary sequences largely untouched.

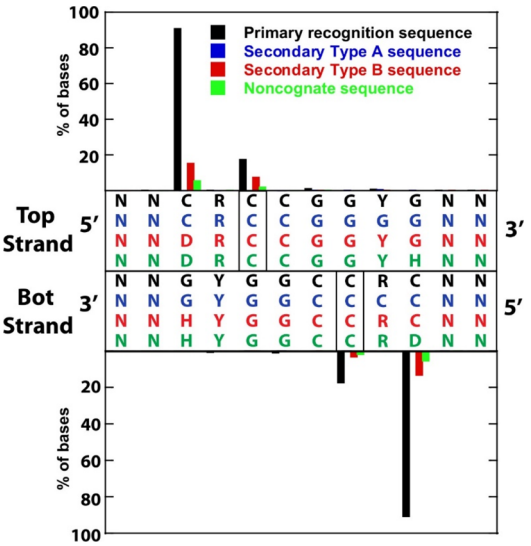


Figure 5. PacBio data of *S. griseus* genomic DNA. Sequences at each position are shown below or above the charted data for the two strands (Y = C or T, R = A or G, D = G, A, or T, H = C, A, or T). The *y* axis plots the percentage of the given base in the respective sequence, which provides an IPD signal >2 (IPD, interpulse duration). Methylation is indicated by a signal above background at the site as well as at the position two nucleotides away in the 5′ direction. Only the cognate sequences show significant methylation at the expected site: CR**C**^{m}CGGYG.

DISCUSSION

Possible Structural Origin of Increased L State Stability when Secondary Sequences Are Bound. The current kinetic study was undertaken to test the hypothesis that the addition of a single step describing an equilibrium between low and high activity conformations of DNA bound SgrAI (i.e., between L and H states) could adapt our prior computational model developed for cleavage of primary site DNA to the cleavage of secondary site DNA by SgrAI. The modeling was successful in that the simulated data matched the experimental data very well (Table S1 and Figures S6 and S8), and the best-fit value for the rate constant for the conversion of the H form to L was found to be 5.9 s^{-1}. Since the L to H rate constant was held at 1 s^{-1}, the fitting indicates that the L state is stabilized ∼6-fold over the H. Because rate constants derived from fitting primary site cleavage data were used in all other steps of the reaction mechanism and this L to H step was not used in fitting that data, the ∼6-fold stability of the L over H state for SgrAI bound to the secondary sequence is relative that for SgrAI bound to the primary sequence. However, the length of the reporter DNA used to measured DNA cleavage rates also differs in the two studies. The prior study with the primary sequence was done with an 18 bp reporter DNA, while the current study was performed with a secondary sequence embedded in a 26 bp reporter DNA. It was necessary to increase the length of the reporter DNA to observe robust activation of DNA cleavage; a secondary sequence embedded in 18 bp shows only minimal activation. These differ only in

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

the number of bp flanking the recognition sequence, with 9 flanking bp in the 26 bp construct and 5 in the 18 bp DNA. The flanking DNA is important for making stabilizing contacts between DBD in the filament by as much as sevenfold (see Results). The combination of the sevenfold and sixfold differences estimate that the L state may be favored by up to ∼40-fold when the secondary sequence is bound, compared to primary.

It should be noted that the L state is favored for both types of complexes, with SgrAI bound to primary or to secondary sequences. The degree to which the L state is favored over the H by SgrAI bound to *primary* sequences can be estimated if it is assumed that (1) only the H state is capable of cleaving DNA, (2) the rate constant for DNA cleavage by SgrAI in the filamentous state of 0.8 s^{-1} represents the true intrinsic cleavage rate constant of the H state, and (3) the cleavage rate constant in the absence of filamentation (0.0009 s^{-1}) provides an estimate of the frequency or proportion of SgrAI in the L and H states. Hence, the ratio of 9×10^{-4} s^{-1}/0.8 s^{-1} suggests that the H state is occupied in only 1 out of 1000 copies of SgrAI (and that the L state is favored by 1000-fold over the H) in the absence of filamentation when the primary sequence is bound. This analysis also suggests that the very slow rate constant for the addition of DBD to filaments is due to the 1000-fold (in the case of the primary sequence) and 4×10^4-fold (in the case of the secondary) preference of the L state by SgrAI bound to DNA, meaning that most collisions between DBD occur between filamentation incompetent (i.e., L state) species and are therefore nonproductive in terms of filament assembly.

To identify how a single base pair substitution in the recognition sequence could result in an up to 40-fold preference for the low activity L conformation by SgrAI, structures of SgrAI in both L and H states were compared.[34,36,42,43] In the case of the primary sequence, structures are available in both the L and H states; however, only an L state structure is available for SgrAI bound to a secondary sequence.[43] Comparing the L state structures of SgrAI bound to primary and to secondary sequences show no significant differences beyond the simple base substitution.[43] SgrAI recognizes the second (and seventh) base pair of its recognition sequences solely via indirect readout; hence, the base substitutions at the second bp did not affect hydrogen bonds or other direct readout contacts with SgrAI.[42]

Since no structures of SgrAI bound to a secondary sequence DNA in the H state are known, it was necessary to resort to comparing the available H and L structures of SgrAI bound to a primary sequence. Many differences in conformation in both SgrAI and the bound DNA are observed, but most likely to affect the SgrAI bound to either type of sequence equally. One structural difference was identified, which would impact complexes with secondary sequences substituted in the first bp, and another that would impact those with the substitution in the second position (as used in this study).[34,36] This latter change occurs in the DNA and results in greater stacking of the second and third base pairs of the recognition sequence in the H conformation compared to the L (Figure 1E).[34] Because the identity of the second nucleotide differs in the two types of sequences (C in the secondary sequence and A or G in the primary sequence), the energy associated with the difference in stacking will also differ. Estimates of base stacking energy indicate that as much as −0.6 kcal/mol (at 25 °C) more could be achieved in the H state with the primary sequence compared to secondary;[34,58] stated in the converse, the L state is more favored than the H by −0.6 kcal/mol with secondary compared to when primary sequences are bound. In this report, a ∼40-fold preference for the preference for the L state over the H with the secondary sequence (compared to that with the primary) was found, which translates to −2.2 kcal/mol, much larger than current estimates that the role DNA structure could play in preferential L state stabilization. It should be noted that estimates of base stacking energy are measured with model sequences and the precise DNA structure is not known.[58] Furthermore, structural comparisons between the L and H states were done with SgrAI bound to only primary sequences. The structure of the H state of SgrAI bound to the secondary sequence will be important to fully understand the larger L state preference for SgrAI bound to this type of sequence.

The SgrAI Filamentation Mechanism and the Biological Role of SgrAI. In the experimental setting, DNA oligonucleotides containing only a single SgrAI recognition site, and at relatively high concentrations, are used to measure DNA cleavage rates, whereas in the natural setting of the host cell, recognition sites would be found on the same contiguous DNA and concentrations of DNA molecules are low (i.e., one copy of the host genome per cell). Primary recognition sites on the host genome (*S. griseus*) are methylated and hence neither bound nor cleaved by SgrAI. This current work now shows conclusively that secondary sequences in the host genome are not methylated and are therefore available for binding to SgrAI (SgrAI binds both primary and secondary sequences with low nanomolar affinity[37]). However, cleavage of these unmethylated secondary sequences is not expected to occur without the availability of unmethylated primary sequences to induce filament formation by SgrAI.

When invading DNA such as phages encounter SgrAI in the cell, the unmethylated primary sequences will induce filamentation by SgrAI, which will then draw in SgrAI bound to secondary sequences. This poses a risk for secondary sequences in the host genome. It is the slow filament assembly step that prevents SgrAI bound to sites on the host genome from being drawn into the filament. Because the rate constant for this step is so low (3×10^{-4} nM^{-1} s^{-1}), only high concentrations of SgrAI bound to DNA can produce biologically relevant rates for filament assembly (and therefore DNA cleavage). The concentration of DNA in the cell is very low, with one to two molecules per cell (host genome and invading DNA), and SgrAI on different DNA molecules are estimated at 1 nM.[22] The rate of association of these two DNA bound SgrAI into a filament is then predicted to be (3×10^{-4} nM^{-1} s^{-1})(1 nM)(1 nM) or 3×10^{-4} nM/s.[22] Meanwhile, SgrAI bound to sequences on the same DNA experience higher "local" concentrations, which can be estimated by considering the approximate distance between the recognition sequences along the contiguous DNA. For example, the *Streptomyces* phage ΦC31 has 2 primary and 55 secondary in a genome of 40 kb, giving an average distance between adjacent primary sequences of 6 kb and between adjacent primary and secondary sequences of 0.5 kb. However, DNA molecules tend to coil rather than form extended structures, and as a result, sequences will be closer in space than predicted by the linear distance between them along the DNA chain. To estimate this distance, theory developed by von Hippel and coworkers predicts a radius of gyration to describe the average distance between two sites on the same DNA molecule given

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

the number of base pairs between them and a DNA persistence length of 500 Å.[49] Using this equation, the local concentration of SgrAI bound to two primary sequences on the same DNA molecule was previously estimated to be as high as 80 nM and between primary and secondary to be as high as 700 nM.[22] With these concentrations, the rates of association of two SgrAI bound to primary sequences on the same DNA molecule into filaments could be as high as $(3 \times 10^{-4}\ nM^{-1}\ s^{-1})(80\ nM)(80\ nM)$ or ∼2 nM/s and between SgrAI bound to primary and to secondary as high as $(3 \times 10^{-4}\ nM^{-1}\ s^{-1})(1/40)(700\ nM)(700\ nM)$ or ∼4 nM/s. These rates are four orders of magnitude faster than for SgrAI bound to sites on separate DNA molecules[22] and suggests that the filament mechanism may have evolved to rapidly cleave the additional secondary sequences in phage DNA without resulting in harming DNA cleavage of the host genome (Figure 6).

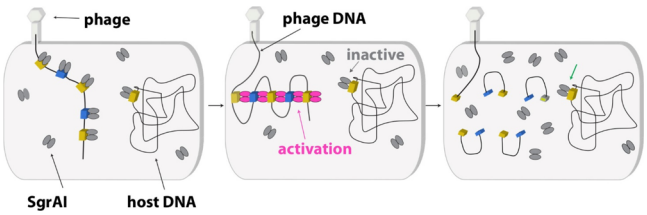


Figure 6. Schematic of SgrAI filamentation mechanism in the context of the cell. Left: SgrAI (gray) binds to primary (blue) and secondary (gold) recognition sites in the host genome and invading phage DNA. Primary sites are methylated and therefore unavailable for SgrAI binding in the host genome. The host genome contains over 1000 secondary sequences, but for simplicity, only one is shown. Phage genomes typically contain two or more primary sequences and 10 or more secondary sequences. Middle: SgrAI bound to primary sequences on the phage DNA nucleate filaments that incorporate SgrAI bound to secondary sequences but only those on the phage DNA. Those on the host genome are not drawn into the filament. Right: primary and secondary sequences on the phage DNA are cleaved, leaving the host genome untouched (green arrow).

Why the Two Types of Recognition Sequences? We have previously speculated that the unique attributes of the SgrAI system evolved as a result of the much larger genome of its *Streptomyces* host. The larger host genome (8 Mbps, vs that of *E. coli* of 5 Mbps) results in many more primary recognition sequences, which require protection via methylation by the cognate SgrAI methyltransferase lest they be cleaved by the SgrAI endonuclease. Methyltransferases utilize the cofactor S-adenosylmethionine (SAM), which is made at some expense by the cell. The longer recognition sequences of SgrAI (8 vs 4−6 bp) may have evolved to reduce the number of recognition sites and therefore reduce the burden on the methyltransferase. In addition, the lower DNA cleavage rate of SgrAI (prior to activation) of 0.0009 s^{-1} (compared to 0.1 s^{-1} for the HincII enzyme[59]) should also reduce the chances of host genome cleavage prior to methylation by SgrAI.M. The recognition sequence of the SgrAI enzyme is relatively long compared to other type II restriction endonucleases (8 bp vs the typical 4−6 bp). However, the longer recognition sequence will also occur less frequently in invading phage DNA and, combined with the slower DNA cleavage rate of SgrAI, could allow the phage to escape by initiating transcription and replication and/or via methylation by SgrAI.M (and thereby protecting it from the SgrAI endonuclease). The filament-forming mechanism of SgrAI solves both of these problems. First, SgrAI is activated to rates comparable to other type II restriction endonucleases, but only when filamentation is induced by the invading DNA. Second, the secondary sequence cleavage activity increases the number of cleavage sites in phage (from 3 possible recognition sequences to 17), which in turn are predicted to prevent phage replication more effectively.[60−62] Hence, the damaging DNA cleavage activity of SgrAI is sequestered on the invading DNA, leaving the host DNA untouched (Figure 6).

CONCLUSIONS

The studies herein add to our growing understanding of the ways enzyme filamentation can control enzymatic activity, in terms of both catalytic rate and substrate specificity. SgrAI is one of a growing list of filament-forming enzymes, many of which have yet to be fully characterized for the effect that filamentation has on enzyme activity and cellular function. Future studies on the filament-forming mechanism of SgrAI will be required to investigate the details of the filamentation induced change in DNA sequence specificity, such as the determination of the H state (i.e., filamented) structure of SgrAI bound to the secondary sequence used in this study. Finally, studies with the secondary sequences substituted in the first bp (i.e., DRCCGGYG, D = A, G, or T) will also be important to determine how filamentation is modulated by this class of secondary sequence, which may utilize different mechanisms such as a disorder-to-order transition to fine-tune the conformational energy landscape of SgrAI.[36]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.3c00313.

(Table S1) Experimental data including SgrAI and DNA concentrations, goodness of fit between simulated and experimental data (R^2), and single exponential constants for analytical fitting; (Figure S1) fluorescence emission from uncleaved reporter single and double-stranded DNAs; (Figure S2) fluorescence of "precleaved" single-stranded reporter DNAs; (Figure S3) changes in fluorescence of reporter DNA with the addition of SgrAI, activator (PC DNA), $MgCl_2$, and SDS; (Figure S4) equilibrium binding measurements of precleaved reporter DNA; (Figure S5) Fitspace analysis of the L/H equilibrium vs the apparent rate constant for SgrAI binding to the cleaved Flo-26M-2-Rox DNA; (Figure S6) fluorescence unquenching and simulated curves from global data fitting to the Kintek model; (Figure S7) representative scans of denaturing PAGE gels used to quantitate DNA cleavage; (Figure S8) experimental DNA cleavage data measured using denaturing gel electrophoresis and simulated curves from global data fitting; (Figure S9) rate constants for each data set determined using a single exponential equation; and (Table S2) Kintek Global Kinetic Explorer Equations (PDF)

Related Articles

The genome sequence of *Streptomyces griseus* NEB1061 along with the raw PacBio reads have been deposited into NCBI BioProject number PRJNA1041797.

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

AUTHOR INFORMATION

Corresponding Author

Nancy C. Horton − *Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, United States;* orcid.org/0000-0003-2710-8284; Email: nhorton@u.arizona.edu

Authors

Niloofar Ghadirian − *Department of Chemistry & Biochemistry, University of Arizona, Tucson, Arizona 85721, United States;* Present Address: Present Address: Department of Molecular Pathobiology, New York University, 433 first Avenue, New York, New York 10010, United States (N.G.)

Richard D. Morgan − *New England Biolabs, Inc., Ipswich, Massachusetts 01938, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.3c00313

Funding

Research reported in this publication was supported by the National Science Foundation under Grant No. MCB-1934291 (to N.C.H.) and MRI DBI-2018942 (to N.C.H.) and the University of Arizona Research, Innovation & Impact (RII) and Technology Research Initiative Fund/Improving Health and Access and Workforce Development. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NSF.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Beaty, N. B.; Lane, M. D. Kinetics of activation of acetyl-CoA carboxylase by citrate. Relationship to the rate of polymerization of the enzyme. *J. Biol. Chem.* **1983**, *258*, 13043−13050.

(2) Meredith, M. J.; Lane, M. D. Acetyl-CoA carboxylase. Evidence for polymeric filament to protomer transition in the intact avian liver cell. *J. Biol. Chem.* **1978**, *253*, 3381−3383.

(3) Kleinschmidt, A. K.; Moss, J.; Lane, D. M. Acetyl coenzyme A carboxylase: filamentous nature of the animal enzymes. *Science* **1969**, *166*, 1276−1278.

(4) Frey, T. G.; Eisenberg, D.; Eiserling, F. A. Glutamine synthetase forms three- and seven-stranded helical cables. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 3402−3406.

(5) Gunning, B. E. The Fine Structure of Chloroplast Stroma Following Aldehyde Osmium-Tetroxide Fixation. *J. Cell Biol.* **1965**, *24*, 79−93.

(6) Kessler, D.; Herth, W.; Knappe, J. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of Escherichia coli. *J. Biol. Chem.* **1992**, *267*, 18073−18079.

(7) Kessler, D.; Leibrecht, I.; Knappe, J. Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of Escherichia coli reside on a polymeric protein particle encoded by adhE. *FEBS Lett.* **1991**, *281*, 59−63.

(8) Josephs, R.; Borisy, G. Self-assembly of glutamic dehydrogenase into ordered superstructures: multichain tubes formed by association of single molecules. *J. Mol. Biol.* **1972**, *65*, 127−155.

(9) Angermuller, S.; Bruder, G.; Volkl, A.; Wesch, H.; Fahimi, H. D. Localization of xanthine oxidase in crystalline cores of peroxisomes. A cytochemical and biochemical study. *Eur. J. Cell Biol.* **1987**, *45*, 137−144.

(10) Olsen, B. R.; Svenneby, G.; Kvamme, E.; Tveit, B.; Eskeland, T. Formation and ultrastructure of enzymically active polymers of pig renal glutaminase. *J. Mol. Biol.* **1970**, *52*, 239−245.

(11) Antonini, E.; Brunori, M.; Bruzzesi, R.; Chiancone, E.; Massey, V. Association-dissociation phenomena of D-amino acid oxidase. *J. Biol. Chem.* **1966**, *241*, 2358−2366.

(12) Kemp, R. G. Rabbit liver phosphofructokinase. Comparison of some properties with those of muscle phosphofructokinase. *J. Biol. Chem.* **1971**, *246*, 245−252.

(13) Trujillo, J. L.; Deal, W. C., Jr. Pig liver phosphofructokinase: asymmetry properties, proof of rapid association–dissociation equilibria, and effect of temperature and protein concentration on the equilibria. *Biochemistry* **1977**, *16*, 3098−3104.

(14) Reinhart, G. D.; Lardy, H. A. Rat liver phosphofructokinase: kinetic and physiological ramifications of the aggregation behavior. *Biochemistry* **1980**, *19*, 1491−1495.

(15) Narayanaswamy, R.; Levy, M.; Tsechansky, M.; Stovall, G. M.; O'Connell, J. D.; Mirrielees, J.; Ellington, A. D.; Marcotte, E. M. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10147−10152.

(16) Werner, J. N.; Chen, E. Y.; Guberman, J. M.; Zippilli, A. R.; Irgon, J. J.; Gitai, Z. Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7858−7863.

(17) Noree, C.; Sato, B. K.; Broyer, R. M.; Wilhelm, J. E. Identification of novel filament-forming proteins in Saccharomyces cerevisiae and Drosophila melanogaster. *J. Cell Biol.* **2010**, *190*, 541−551.

(18) Liu, J. L. Intracellular compartmentation of CTP synthase in Drosophila. *J. Genet Genomics* **2010**, *37*, 281−296.

(19) Suresh, H. G.; da Silveira Dos Santos, A. X.; Kukulski, W.; Tyedmers, J.; Riezman, H.; Bukau, B.; Mogk, A. Prolonged starvation drives reversible sequestration of lipid biosynthetic enzymes and organelle reorganization in Saccharomyces cerevisiae. *Molecular biology of the cell* **2015**, *26*, 1601−1615.

(20) Shen, Q. J.; Kassim, H.; Huang, Y.; Li, H.; Zhang, J.; Li, G.; Wang, P. Y.; Yan, J.; Ye, F.; Liu, J. L. filamentation of Metabolic Enzymes in *Saccharomyces cerevisiae*. *J. Genet. Genom.* **2016**, *43*, 393−404.

(21) Park, C. K.; Horton, N. C. Structures, functions, and mechanisms of filament forming enzymes: a renaissance of enzyme filamentation. *Biophys Rev.* **2019**, *11*, 927−994.

(22) Barahona, C. J.; Basantes, L. E.; Tompkins, K. J.; Heitman, D. M.; Chukwu, B. I.; Sanchez, J.; Sanchez, J. L.; Ghadirian, N.; Park, C. K.; Horton, N. C.; Shisler, J. L. The need for speed: run-on oligomer filament formation provides maximum speed with maximum sequestration of activity. *J. Virol* **2019**, *93*, 1−19.

(23) Johnson, M. C.; Kollman, J. M. Cryo-EM structures demonstrate human IMPDH2 filament assembly tunes allosteric regulation. *Elife* **2020**, *9*, No. e53243.

(24) Woodward, J. D.; Trompetter, I.; Sewell, B. T.; Piotrowski, M. Substrate specificity of plant nitrilase complexes is affected by their helical twist. *Commun. Biol.* **2018**, *1*, 186.

(25) Hunkeler, M.; Hagmann, A.; Stuttfeld, E.; Chami, M.; Guri, Y.; Stahlberg, H.; Maier, T. Structural basis for regulation of human acetyl-CoA carboxylase. *Nature* **2018**, *558*, 470−474.

(26) Webb, B. A.; Dosey, A. M.; Wittmann, T.; Kollman, J. M.; Barber, D. L. The glycolytic enzyme phosphofructokinase-1 assembles into filaments. *J. Cell Biol.* **2017**, *216*, 2305−2313.

(27) Lynch, E. M.; Hicks, D. R.; Shepherd, M.; Endrizzi, J. A.; Maker, A.; Hansen, J. M.; Barry, R. M.; Gitai, Z.; Baldwin, E. P.; Kollman, J. M. Human CTP synthase filament structure reveals the active enzyme conformation. *Nat. Struct Mol. Biol.* **2017**, *24*, 507−514.

(28) Anthony, S. A.; Burrell, A. L.; Johnson, M. C.; Duong-Ly, K. C.; Kuo, Y. M.; Simonet, J. C.; Michener, P.; Andrews, A.; Kollman, J. M.; Peterson, J. R. Reconstituted IMPDH polymers accommodate both catalytically active and inactive conformations. *Molecular biology of the cell* **2017**, *28*, 2600−2608.

(29) Korennykh, A. V.; Egea, P. F.; Korostelev, A. A.; Finer-Moore, J.; Zhang, C.; Shokat, K. M.; Stroud, R. M.; Walter, P. The unfolded

protein response signals through high-order assembly of Ire1. *Nature* **2009**, *457*, 687−693.
(30) Zhong, J.; Guo, C. J.; Zhou, X.; Chang, C. C.; Yin, B.; Zhang, T.; Hu, H. H.; Lu, G. M.; Liu, J. L. Structural basis of dynamic P5CS filaments. *Elife* **2022**, *11*, No. e76107.
(31) Bennett, J. A.; Steward, L. R.; Rudolph, J.; Voss, A. P.; Aydin, H. The structure of the human LACTB filament reveals the mechanisms of assembly and membrane binding. *PLoS biology* **2022**, *20*, No. e3001899.
(32) Zhang, M.; Zhang, L.; Guo, R.; Xiao, C.; Yin, J.; Zhang, S.; Yang, M. Structural basis for the catalytic activity of filamentous human serine beta-lactamase-like protein LACTB. *Structure* **2022**, *30* (685−696), No. e685.
(33) Huang, P.; Chen, S.; Chiang, W.; Ho, M.; Wu, K. Structural basis for the helical filament formation of Escherichia coli glutamine synthetase. *Protein Sci.* **2022**, *31*, No. e4304.
(34) Polley, S.; Lyumkis, D.; Horton, N. C. Mechanism of filamentation-induced allosteric activation of the SgrAI endonuclease. *Structure* **2019**, *27*, 1497−1507.
(35) Lyumkis, D.; Talley, H.; Stewart, A.; Shah, S.; Park, C. K.; Tama, F.; Potter, C. S.; Carragher, B.; Horton, N. C. Allosteric regulation of DNA cleavage and sequence-specificity through run-on oligomerization. *Structure* **2013**, *21*, 1848−1858.
(36) Shan, Z.; Ghadirian, N.; Lyumkis, D.; Horton, N. C. Pretransition state and apo structures of the filament-forming enzyme SgrAI elucidate mechanisms of activation and substrate specificity. *J. Biol. Chem.* **2022**, *298*, No. 101760.
(37) Park, C. K.; Stiteler, A. P.; Shah, S.; Ghare, M. I.; Bitinaite, J.; Horton, N. C. Activation of DNA cleavage by oligomerization of DNA-bound SgrAI. *Biochemistry* **2010**, *49*, 8818−8830.
(38) Bitinaite, J.; Schildkraut, I. Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1164−1169.
(39) Shah, S.; Sanchez, J.; Stewart, A.; Piperakis, M. M.; Cosstick, R.; Nichols, C.; Park, C. K.; Ma, X.; Wysocki, V.; Bitinaite, J.; Horton, N. C. Probing the run-on oligomer of activated SgrAI bound to DNA. *PLoS One* **2015**, *10*, No. e0124783.
(40) Bilcock, D. T.; Daniels, L. E.; Bath, A. J.; Halford, S. E. Reactions of type II restriction endonucleases with 8-base pair recognition sites. *J. Biol. Chem.* **1999**, *274*, 36379−36386.
(41) Ma, X.; Shah, S.; Zhou, M.; Park, C. K.; Wysocki, V. H.; Horton, N. C. Structural Analysis of Activated SgrAI-DNA Oligomers Using Ion Mobility Mass Spectrometry. *Biochemistry* **2013**, *52*, 4373−4381.
(42) Dunten, P. W.; Little, E. J.; Gregory, M. T.; Manohar, V. M.; Dalton, M.; Hough, D.; Bitinaite, J.; Horton, N. C. The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res.* **2008**, *36*, 5405−5416.
(43) Little, E. J.; Dunten, P. W.; Bitinaite, J.; Horton, N. C. 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.* **2011**, *67*, 67−74.
(44) Beese, L. S.; Steitz, T. A. Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* **1991**, *10*, 25−33.
(45) Strater, N.; Lipscomb, W. N.; Klabunde, T.; Krebs, B. Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2024−2055.
(46) Yang, W. Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011**, *44*, 1−93.
(47) Park, C. K.; Sanchez, J. L.; Barahona, C.; Basantes, L. E.; Sanchez, J.; Hernandez, C.; Horton, N. C. The run-on oligomer filament enzyme mechanism of SgrAI: Part 1. Assembly kinetics of the run-on oligomer filament. *J. Biol. Chem.* **2018**, *293*, 14585−14598.
(48) Park, C. K.; Sanchez, J. L.; Barahona, C.; Basantes, L. E.; Sanchez, J.; Hernandez, C.; Horton, N. C. The run-on oligomer filament enzyme mechanism of SgrAI: Part 2. Kinetic modeling of the full DNA cleavage pathway. *J. Biol. Chem.* **2018**, *293*, 14599−14615.
(49) Rippe, K.; von Hippel, P. H.; Langowski, J. Action at a distance: DNA-looping and initiation of transcription. *Trends in biochemical sciences* **1995**, *20*, 500−506.
(50) Oehler, S.; Muller-Hill, B. High local concentration: a fundamental strategy of life. *J. Mol. Biol.* **2010**, *395*, 242−253.
(51) Blow, M. J.; Clark, T. A.; Daum, C. G.; Deutschbauer, A. M.; Fomenkov, A.; Fries, R.; Froula, J.; Kang, D. D.; Malmstrom, R. R.; Morgan, R. D.; Posfai, J.; Singh, K.; Visel, A.; Wetmore, K.; Zhao, Z.; Rubin, E. M.; Korlach, J.; Pennacchio, L. A.; Roberts, R. J. The Epigenomic Landscape of Prokaryotes. *PLoS Genet* **2016**, *12*, No. e1005854.
(52) Clark, T. A.; Lu, X.; Luong, K.; Dai, Q.; Boitano, M.; Turner, S. W.; He, C.; Korlach, J. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.* **2013**, *11*, 4.
(53) Fasman, G. D. *CRC Handbook of Biochemistry and Molecular Biology*, 3rd ed., CRC: Cleveland, OH. (1975).
(54) Johnson, K. A.; Simpson, Z. B.; Blom, T. Fitspace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem.* **2009**, *387*, 30−41.
(55) Johnson, K. A.; Simpson, Z. B.; Blom, T. Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **2009**, *387*, 20−29.
(56) Johnson, K. A. Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol* **2009**, *467*, 601−626.
(57) Roberts, R. J.; Vincze, T.; Posfai, J.; Macelis, D. REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **2010**, *38*, D234−236.
(58) Bommarito, S.; Peyret, N.; SantaLuciar, J., Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **2000**, *28*, 1929−1934.
(59) Joshi, H. K.; Etzkorn, C.; Chatwell, L.; Bitinaite, J.; Horton, N. C. Alteration of sequence specificity of the type II restriction endonuclease HincII through an indirect readout mechanism. *J. Biol. Chem.* **2006**, *281*, 23852−23869.
(60) Moineau, S.; Pandian, S.; Klaenhammer, T. R. Restriction/Modification systems and restriction endonucleases are more effective on lactococcal bacteriophages that have emerged recently in the dairy industry. *Appl. Environ. Microbiol.* **1993**, *59*, 197−202.
(61) Lee, S.; Ward, T. J.; Siletzky, R. M.; Kathariou, S. Two novel type II restriction-modification systems occupying genomically equivalent locations on the chromosomes of Listeria monocytogenes strains. *Appl. Environ. Microbiol.* **2012**, *78*, 2623−2630.
(62) Kasarjian, J. K. A. New restriction enzymes discovered from Escherichia coli clinical strains using a plasmid transformation method. *Nucleic Acids Res.* **2003**, *31*, No. e22.

https://doi.org/10.1021/acs.biochem.3c00313
Biochemistry 2024, 63, 326−338

Supporting Information for:

DNA Sequence Control of Enzyme Filamentation and Activation of the SgrAI Endonuclease

Niloofar Ghadirian[1,2], Richard D. Morgan[3] & Nancy C. Horton[4*]

[1]Department of Chemistry & Biochemistry, University of Arizona, Tucson, Arizona, 85721

[2]Present Address: Department of Molecular Pathobiology, New York University, 433 1st Avenue, New York, New York, US, 10010

[3]New England Biolabs, Inc., Ipswich, MA 01938

[4]Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721

***To whom correspondence should be addressed:**

Dr. Nancy C. Horton

Department of Molecular and Cellular Biology

Telephone: (520) 626-3828.

E-mail: nhorton@u.arizona.edu.

Table of Contents

Table S1. Experimental Data including SgrAI and DNA concentrations, goodness of fit between simulated and experimental data (R^2) and single exponential constants for analytical fitting.

Experiment # in KinTek	Flo-26M-2-Rox Concentration (nM)	Double stranded PC DNA Concentration (nM)	SgrAI Concentration (μM)	R^2 of Kintek Simulated Curves to Experimental Data	Fitted Scaling Factors (SF) and Baseline (BL) parameters	Single Exponential Rate Constant (s^{-1}) and R^2 from Analytical Fits
1	100	100	2.5	Rox: 0.97895 Flo: 0.818531	SF=0.86, 0.76 BL=6.2, 22.0	Rox: $1.16x10^{-3}$ (0.9842) Flo: $3.72x10^{-3}$ (0.8381)
2	100	500	2.5	Rox: 0.94409 Flo: 0.92260	SF=0.82, 0.84 BL=13.3, 13.3	Rox: $5.71x10^{-3}$ (0.9310) Flo: $6.73x10^{-3}$ (0.9179)
3	50	100	1.0	0.97846	SF=3.3 BL= -5.9	$1.46x10^{-3}$ (0.9518)
4	10	250	0.25	0.98253	SF=11.3 BL=0	$3.84x10^{-3}$ (0.9852)
5	10	500	0.5	0.97849	SF=11.3 BL=0.56	$4.23x10^{-3}$ (0.9687)
6	50	400	1.0	0.95457	SF=2.3 BL=2.2	$4.50x10^{-3}$ (0.9587)
7	50	300	1.0	0.93485	SF=2.3 BL=6.0	$3.61x10^{-3}$ (0.9529)
8	50	150	1.0	0.98313	SF=2.6 BL= -11.5	$1.78x10^{-3}$ (0.9715)
9	100	450	2.5	0.89670	SF=1.8 BL=54.8	$5.81x10^{-3}$ (0.9060)
10	50	250	1.0	Rox: 0.93128 Flo: 0.93434	SF=1.4, 2.0 BL=20.0, 19.1	Rox: $4.80x10^{-3}$ (0.9321) Flo: $5.81x10^{-3}$ (0.9361)
11	50	150	1.0	Rox: 0.87465 Flo: 0.82235	SF=1.5, 1.7 BS=17.8, 20.4	Rox: $6.49x10^{-3}$ (0.9280) Flo: $5.13x10^{-3}$ (0.8939)
12	50	50	1.0	Rox: 0.93356 Flo: 0.87688	SF=1.9, 1.9 BL=7.8, 11.5	Rox: $1.54x10^{-3}$ (0.9417) Flo: $1.52x10^{-3}$ (0.9360)
13	50	50	1.0	0.98592	SF=2.2 BL= -4.0	$5.78x10^{-4}$ (0.9469)
14	50	75	1.0	0.97905	SF=2.2 BL=2.8	$1.25x10^{-3}$ (0.9860)
15	50	200	1.0	0.98975	SF=2.5 BL=7.6	$2.02x10^{-3}$ (0.9865)
16	50	300	1.0	0.97555	SF=2.5 BL=7.3	$2.80x10^{-3}$ (0.9855)

17	50	25	1.0	0.98729	SF=3.0 BL=5.7	5.21×10^{-4} (0.9469)

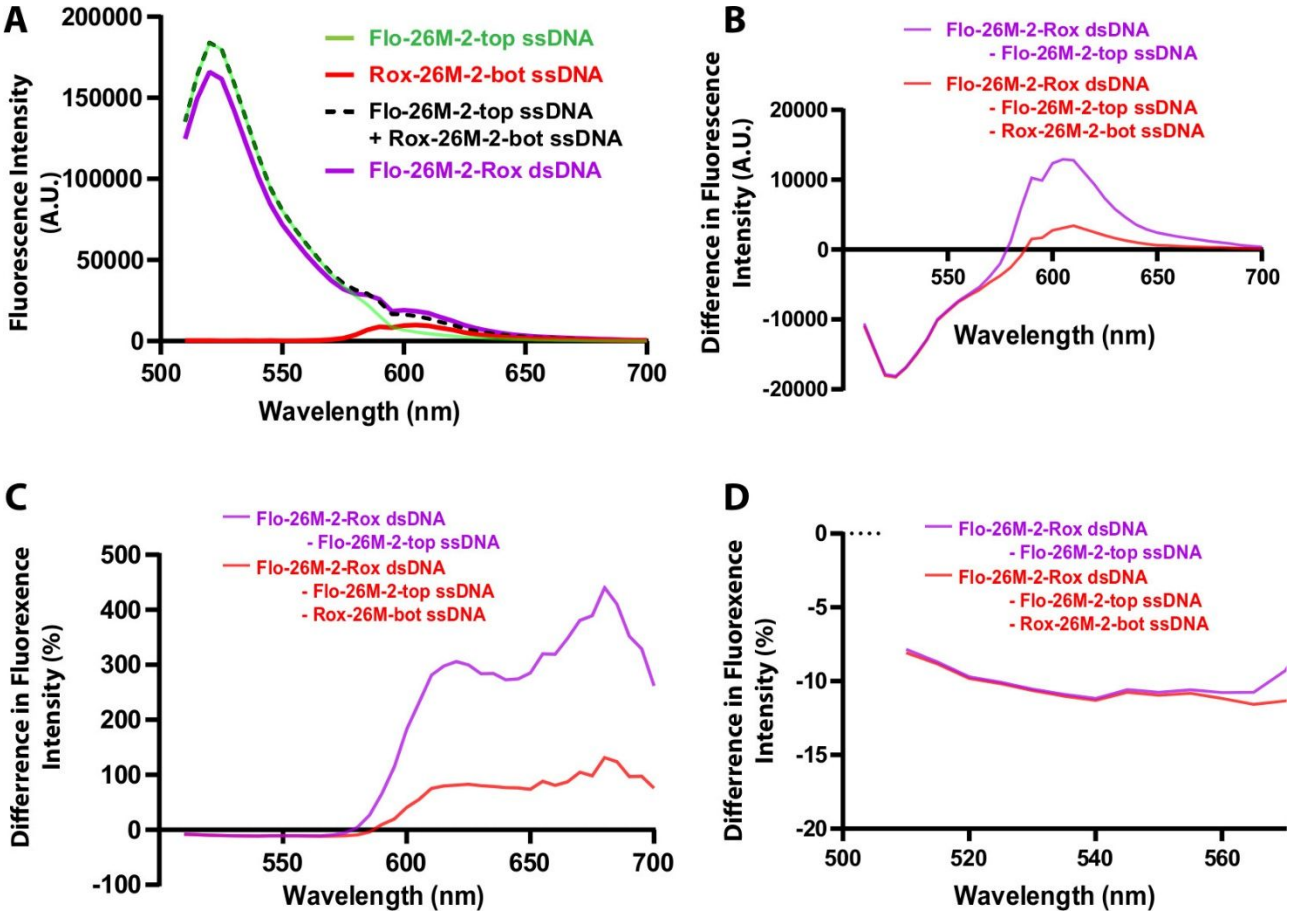


Figure S1. Fluorescence emission from uncleaved reporter single and double-stranded DNAs. A. Fluorescence emission from 50 nM single and double stranded reporter DNAs (ssDNA and dsDNA, respectively) with excitation at 495 nm (i.e. the fluorescein excitation maximum): Flo-26M-2-top ssDNA, green, Rox-26M-2-bot ssDNA, red, the mathematical sum of the Flo-26M-2-top and Rox-26M-2-bot ssDNAs, black dashed line, Flo-26M-2-Rox dsDNA, purple. Measurements were done in 1.5 ml of reaction buffer without $MgCl_2$ (50 mM Tris-HCl, 150 mM NaCl, 10% glycerol, and 1 mM DTT, pH 8.0 at room temperature) and at 25°C. Some quenching of the fluorescein emission (maximum near 520 nm) is seen upon annealing the two strands, as well as a slight increase in Rox emission (maximum near 605 nm)(compare black dashed line to purple). Signals corrected for dilution. **B.** Difference spectra between those shown in A. In purple is the difference between annealed Flo-26M-2-Rox dsDNA and Flo-26M-2-top ssDNA emphasizing quenching of Flo emission at 520 nm and increase in Rox emission at 605 nm upon annealing. The increase in Rox emission

upon annealing appears less when the spectrum for Rox-26M-2-bot ssDNA is further subtracted (red). **C.** As in B, however the y axis is now the percentage difference in emission relative to the spectrum of Flo-26M-2-top ssDNA. This plot shows how the Flo quenching is very little in terms of the percentage of Flo emission (see signal at 520 nm), but the Rox emission resulting from annealing (red) now appears significant in terms of the percentage change at those wavelengths relative to the Flo-26M-2-top ssDNA emission. **D.** As in C, zooming in on the Flo emission maximum wavelengths.

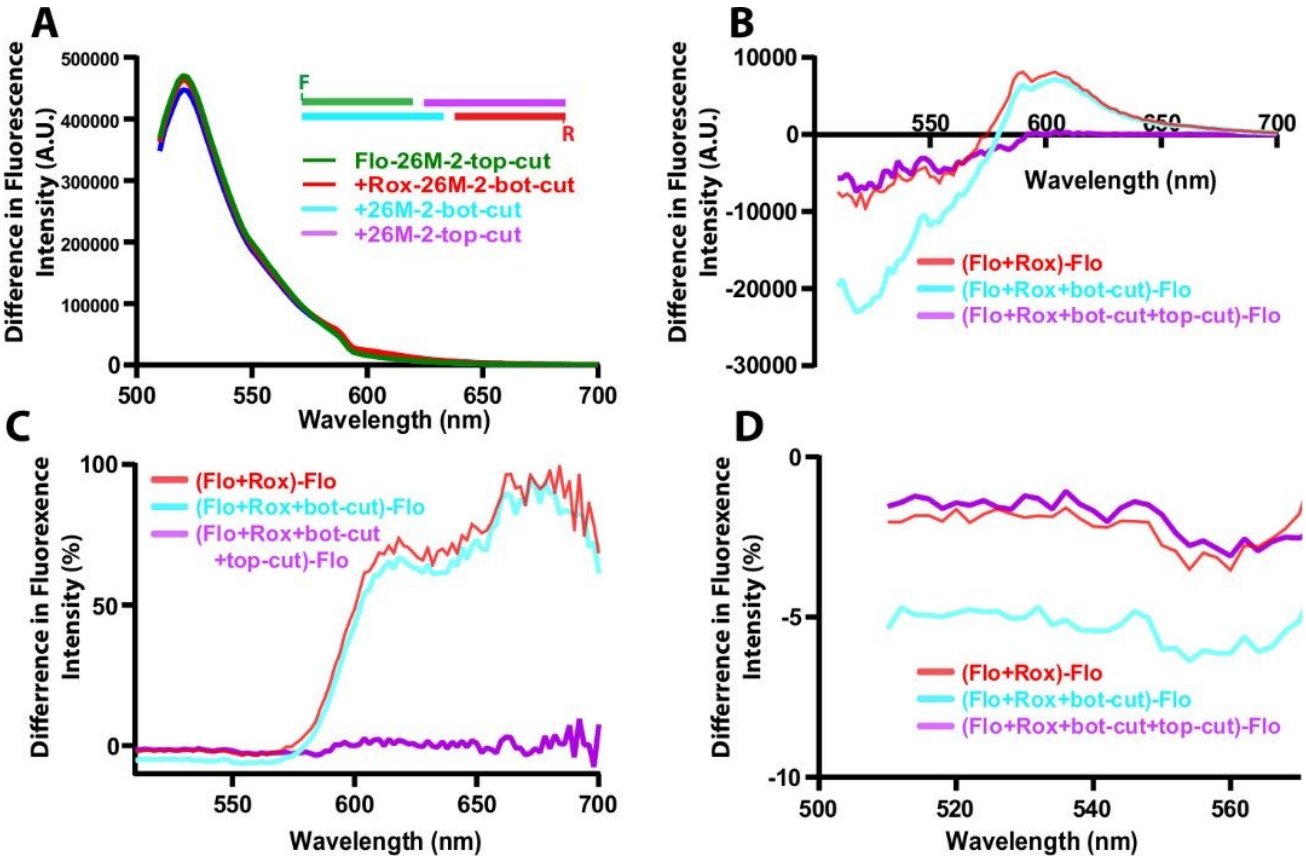


Figure S2. Fluorescence of "pre-cleaved" single stranded reporter DNAs. A. Fluorescence emission (excitation at 495 nm, the fluorescein excitation maximum) of 50 nM Flo-26M-2-top-cut ssDNA only (green), with added 50 nM Rox-26M-2-bot-cut ssDNA (red), further addition of 50 nM 26M-bot-cut (the complement to Flo-26M-2-top cut)(cyan) and finally after further addition of 50 nM 26M-2-top-cut (the complement to Rox-26M-2-bot-cut)(purple). Measurements were done in 1.5 ml of reaction buffer without $MgCl_2$ and at 25°C, and spectra were corrected for dilution. **B.** Difference spectra of spectra in A using that of the Flo-26M-2-top-cut

as the reference spectrum. In red is the spectrum of Flo-26M-2-top-cut subtracted from that of the mixture of Flo-26M-2-top-cut plus Rox-26M-2-bot-cut. Despite the fact that these oligonucleotides should not anneal, some quenching of the Flo emission at 520 nm is seen, as well as an increase in the emission of Rox at 605 nm (likely due to non-FRET emission/absorption). Addition of the complement to Flo-26M-2-top-cut (i.e. 26M-bot-cut in cyan) results in additional Flo quenching which is reversed to some extent by the addition of the complement to Rox-26M-2-bot-cut (purple). The Rox emission at 605 nm is greatest with only the Flo and Rox labeled strands (red). Addition of the complement to the Flo labeled strand has minimal effect on the Rox emission (cyan), however addition of the strand complementary to Rox greatly quenches the Rox emission (purple). **C.** As in B, however expressed as a percent change relative to the Flo-26M-2-top-cut spectrum. The changes in the Flo emission appear minimal compared to the increase in Rox emission (all curves), however even the Rox emission is greatly diminished with the addition of its complementary strand (purple) **D.** As in C, zooming in on the Flo emission wavelengths.

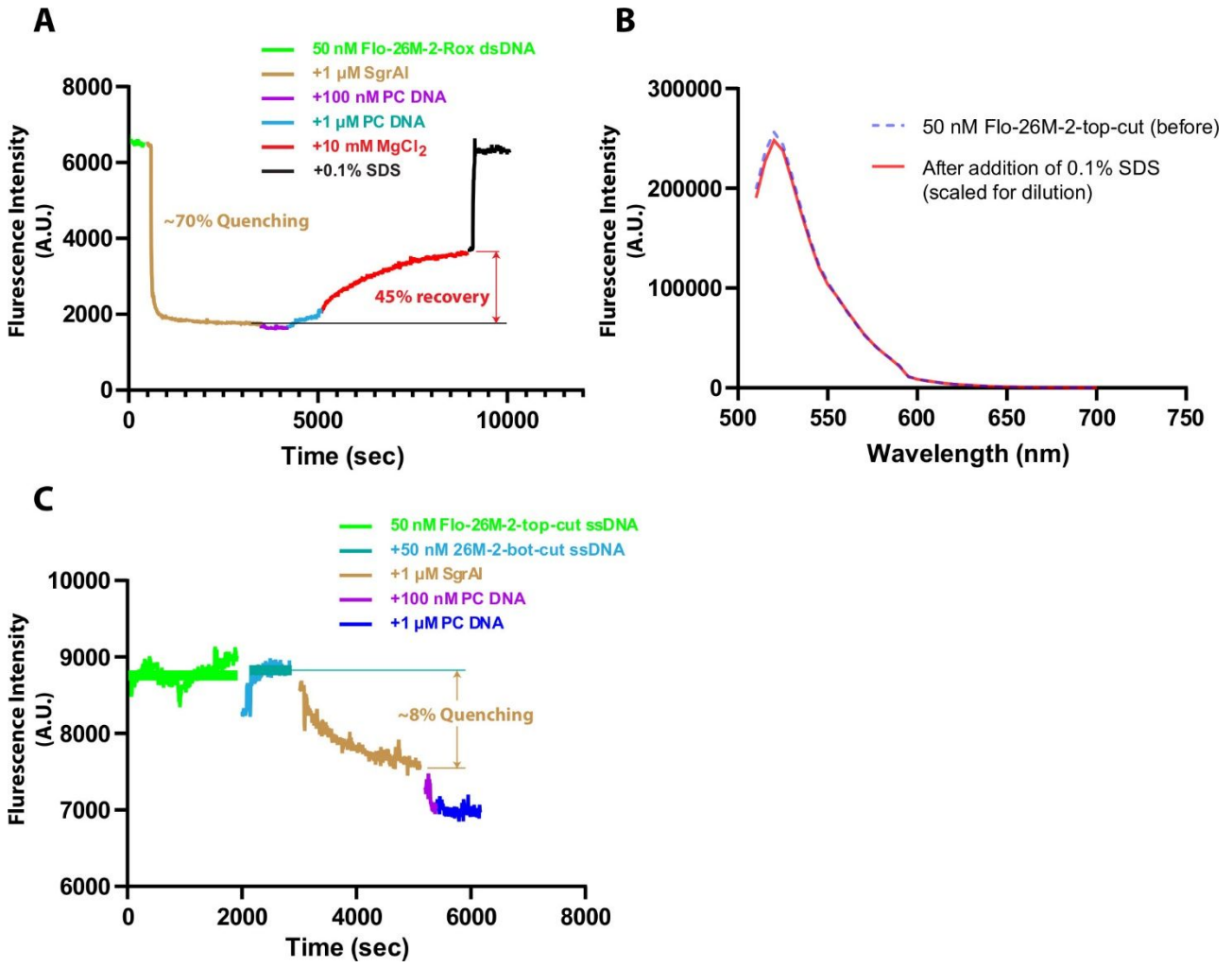


Figure S3. Changes in fluorescence of reporter DNA with the addition of SgrAI, activator (PC DNA), $MgCl_2$, and SDS. A. The fluorescence emission at 520 nm (excitation at 495 nm) of 50 nM Flo-26M-2-Rox dsDNA in 1.5 ml of reaction buffer without $MgCl_2$ and at 25°C is shown in green. The addition of 1 μM SgrAI (gold) results in strong quenching of the fluorescein emission at 520 nm (~70% quenching after correction for dilution). The addition of 100 mM PC DNA and 1 μM PC DNA has minimal effect on the fluorescence (purple and light blue), however the addition of 10 mM $MgCl_2$ (red) results in unquenching via a slow process, presumably due to DNA cleavage and release from SgrAI. The emission does not however reach the original fluorescence (even after considering the effects of dilution which total only 7.5%) until 0.1% (final concentration) SDS is added (black). The fact that SDS is required to recover full fluorescence indicates that some residual binding of SgrAI to the DNA occurs. **B.** Control showing that 0.1% SDS does not affect the

fluorescence of the Flo fluorophore. The emission spectra (with excitation at 495 nm) of 50 nM Flo-26M-2-top-cut in reaction buffer (without Mg^{2+}) at 25°C before (blue dotted line) and after (red solid line) the addition of SDS (spectra corrected to account for dilution). **C.** Fluorescence emission at 520 nm (excitation wavelength of 495 nm) of 50 nM Flo-26M-2-top-cut ssDNA (green) in reaction buffer (with $MgCl_2$). The addition of its complement (50 nM 26M-2-bot-cut ssDNA) does not affect the fluorescence emission (light blue). Addition of 1 μM SgrAI results in a slow progressive quenching (~8% quenching of the initial fluorescence signal after correction for dilution) after ~2000 sec (gold). The addition of 100 nM activator (PC) DNA (purple) results in a small degree of rapid quenching. The addition of 1 μM PC DNA does not result in additional quenching (dark blue). This indicates that SgrAI binds to the "cleaved" DNAs but with low affinity that is dependent on a slow process, likely annealing of the single stranded DNA. Whereas 70% of the uncut Flo labeled DNA was quenched with 1 μM SgrAI, only 8% was quenched with the pre-cleaved version, suggesting that only ~10% binding of the "cleaved" reporter DNAs occurs under these conditions.

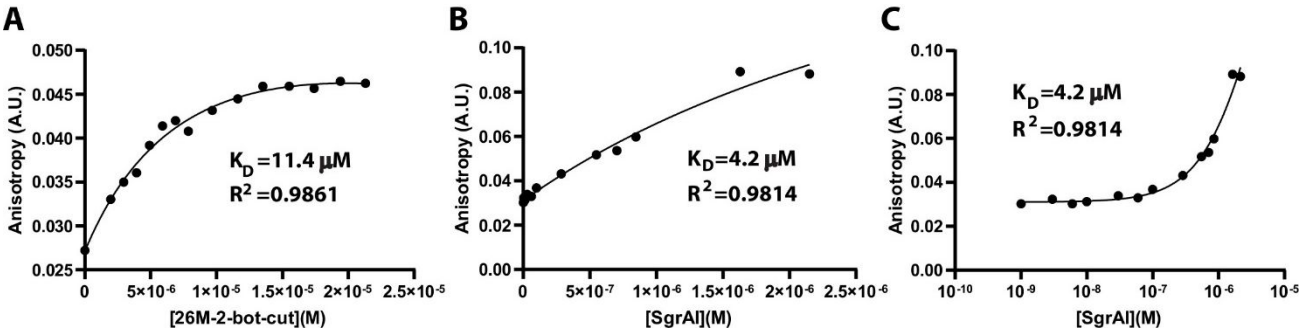


Figure S4. Equilibrium binding measurements of pre-cleaved reporter DNA. A. Titration of 50 nM Flo-26M-2-top-cut ssDNA with its complement 26M-2-bot-cut ssDNA in reaction buffer without $MgCl_2$ at 25°C. Excitation occurred at 495 nm, and emission measured at 520 nm. The fitted equilibrium dissociation constant is 11.4 μM. **B.** Titration of pre-cleaved reporter DNA (50 nM Flo-26M-2-top-cut, 50 nM 26M-2-bot-cut, 50 nM Rox-26M-2-bot-cut, 50 nM 26M-2-top-cut) with SgrAI in reaction buffer without $MgCl_2$. Fitting to an equilibrium binding curve was error prone due to incomplete binding saturating, however the best fit equilibrium dissociation constant was determined to be 4.2 μM. **C.** As in B but with a log X axis.

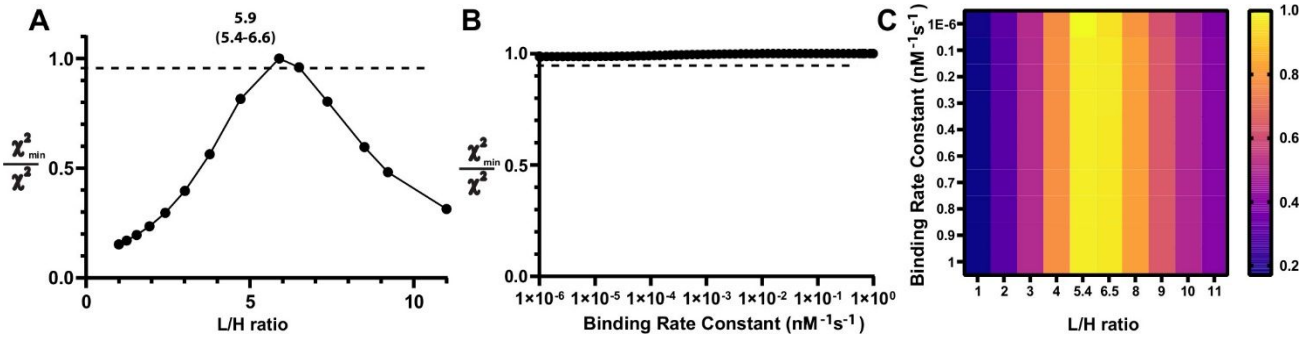


Figure S5. Fitspace analysis of the L/H equilibrium vs. the apparent rate constant for SgrAI binding to the cleaved Flo-26M-2-Rox DNA. A. Fitspace analysis of the quality of fits, measured by χ^2_{min}/χ^2 versus the L/H ratio (i.e. the ratio of rate constants on going from the H to the L state and back; a higher number indicates the degree to which the L state is favored over the H). χ^2_{min}/χ^2 was calculated by dividing the lowest χ^2 found in any fit (i.e. χ^2_{min}) by the χ^2 of the fit with the given value of the L/H ratio, and values closer to 1 indicate the best fits. The quality of the fit depended significantly on the L/H ratio, with maximum occurring at a value of 5.9, with the range of 5.4-6.6 for resulting in 5% of the best χ^2 (dotted line demarcates a χ^2_{min}/χ^2 of 0.95). **B.** As in A, but for the forward rate constant of binding of SgrAI to cleaved Flo-26M-2-Rox. The quality of the fit was insensitive to the value of this rate constant throughout a broad range of values from $1.0\text{x}10^{-6}$ s^{-1} to 1 s^{-1}. **C.** The results of a 2D Fitspace calculation varying the L/H ratio (between 1 and 11) and forward (binding) rate constant of SgrAI to cleaved Flo-26M-2-Rox (between $1\text{x}10^{-6}$ $nM^{-1}s^{-1}$ to 1 $nM^{-1}s^{-1}$). The results show the best fit with the given values for these parameters, in terms of χ^2_{min}/χ^2. During the analysis, all other rate constants were held constant but scaling factors were allowed to be adjusted to give the lowest χ^2. Lighter colors, with χ^2_{min}/χ^2 closer to 1, show the best fits of the model to the data. This analysis showed that the quality of the fit was insensitive to the forward rate constant for binding of SgrAI to the cleaved DNA (y axis) across this range of values, and the best value for the L/H ratio is between 5.4-6.6 s^{-1} (resulting in χ^2 that is within 5% of the best χ^2).

Experiments to determine the origin of Flo quenching and unquenching during single turnover DNA cleavage measurements

The large quenching and incomplete unquenching of the Flo fluorescence emission during DNA cleavage reactions with Flo-26M-2-Rox (the reporter DNA) raised the question of whether other processes were affecting Flo emission in addition to (or in place of) changes in the FRET signal. To investigate the origin of the change in Flo emission during reactions, a series of control studies were performed. First, the effect of annealing the Flo and Rox strands of the reporter DNA (i.e. Flo-26M-2-Rox) was investigated to discover the degree of FRET in this construct. **Figure S1A** shows the emission spectra of single and double-stranded versions of the reporter DNA when excited at 495 nm, the excitation maximum of fluorescein (Flo). These were obtained with 50 nM of the DNAs in 1.5 ml of reaction buffer without Mg^{2+} (50 mM Tris-HCl, 150 mM NaCl, 10% glycerol, and 1 mM DTT, pH 8.0 at room temperature), and at 25°C to mimic the reaction conditions used during DNA cleavage measurements. First, the single stranded DNA containing Flo at the 5' end (Flo-26M-2-top) shows the expected maximum at the Flo emission wavelength (520 nm, solid green line). Rhodamine-X (Rox) has excitation and emission maxima at 568 nm and 605 nm, respectively, and although the excitation wavelength used in **Figure S1A** is 495 nm, the single stranded Rox-26M-2-bot (alone in solution) showed some fluorescence emission (solid red line, **Fig. S1A**) which must be due to a low level of absorbance at 495 nm. The mathematical sum of the spectra of these two independent (i.e. unannealed) single strands is shown as a dashed black line in **Fig. S1A**. The spectra when mixed and annealed (i.e. Flo-26M-2-Rox dsDNA, purple, **Fig. S1A**) shows some quenching of the Flo emission at 520 nm, and a slight increase in emission from Rox at 605 nm, relative to the mathematical sum of the spectra from the two single strands alone (black dashed line, **Fig. S1A**). **Figure S1B** shows the difference spectra of the annealed Flo-26M-2-Rox dsDNA compared to that of the single stranded Flo-26M-2-top ssDNA (purple, **Fig. S1B**), as well as in the spectrum of both single strands subtracted (red, **Fig. S1B**). The difference spectra show the quenching of the Flo emission (at 520 nm, both purple and red lines, **Fig. S1B**), and the increased emission of Rox (at 605 nm), however after subtracting the emission of the single stranded Rox-26M-2-bot, this increase appears to be very much less (compare red to purple lines, **Fig. S1B**). In terms of the percent difference (relative to the spectra of Flo-26M-2-top ssDNA), the quenching of the Flo emission appears minimal (at 520 nm, **Fig. S1C-D**) at about 10-15%. However, by this measure, the Rox emission appears significant, even when the single stranded Rox-26M-2-bot is subtracted (red, **Fig. S1C**). This

is largely due to the fact that very little emission occurs in the reference spectrum (namely that of Flo-26M-2-top ssDNA) at the Rox emission wavelengths (see green line, **Fig. S1A**). The small amount of Rox emission in the duplexed DNA (red in **Fig. S1B** and **S1C**) indicates that a small amount of FRET or non-FRET transfer via absorption by Rox of the Flo emission occurs when the strands are present in the same sample, and which amounts to no more than 10% of the fluorescein signal (**Fig. S1D**).

Next, annealing of the "cleaved" strands of the fluorescence reporter DNA was investigated. It should be noted that these "cleaved" versions are synthetic and therefore do not contain 5'phosphate groups, including that which would naturally occur at the cleavage site. However, prior studies have indicated that this phosphate adds little to the affinity of SgrAI for such "cleaved" constructs[1]. In addition, prior studies have also shown a minimal effect of Mg^{2+} on the measured affinity of SgrAI for its recognition sequence (the K_D of SgrAI for the 40 bp "pre-cleaved" PC DNA is 0.06 nM, and for the secondary sequence embedded in an 18 bp duplex is 2.6 nM)[1]. **Figure S2A** shows the spectra of Flo-26M-2-top-cut ssDNA (green) alone, and after the addition of Rox-26M-2-bot-cut (red), 26M-2-bot-cut (cyan), and 26M-2-top-cut (purple) in reaction buffer at 25°C with excitation at 495 nm. The inset in **Figure S2A** shows how these strands anneal: 26M-2-bot-cut (cyan) is the complement to Flo-26M-2-top-cut (green), and 26M-2-top-cut (magenta) is the complement to Rox-26M-2-bot-cut (red). **Figure S2B** more clearly shows the differences between the spectra. In red is the difference spectra between the emission of the mixture of Flo-26M-2-top-cut and Rox-26M-2-bot-cut, which should not anneal, relative to the spectrum of Flo-26M-2-top-cut alone. Some quenching of the Flo emission (at 550 nm) occurs as well as an increase in emission of the Rox (605 nm). Since these single strands are not expected to anneal (they are not complementary, see sequences in Materials and Methods), this energy transfer must occur through non-FRET radiative means (i.e. absorption by Rox of the Flo emission in the sample). The addition of the complement to Flo-26M-2-top (26M-2-bot-cut) is shown in cyan, and resulted in further quenching of the Flo emission at 520, with little effect on the Rox emission at 605 nm. Some annealing of these DNA strands may be responsible for the observed quenching of the Flo emission. Further addition of the complement to Rox-26M-2-bot-cut, shown in magenta, unquenches the Flo emission (at 520 nm) to some extent while completely quenching the Rox emission at 605 nm. It appears that duplex formation may result in some quenching of both

Flo and Rox emissions, though duplexation of these DNAs at 50 nM concentration is expected to be minimal (see below). **Figures S2C-D** show the difference spectra in terms of percent change from the spectrum of Flo-26M-2-top-cut. The changes in Flo emission at 520 nm seem minor viewed this way (between 3-7%). The increase in Rox emission (red and cyan) appears significant because the emission of Flo-26M-2-top-cut is very low at these wavelengths, but the addition of the complement to the Rox strand eliminates the Rox emission (magenta, **Fig. S2C**). Hence, the presence of the complementary strands appears to quench the Flo emission to a small degree, the Rox emission to perhaps a larger degree, and some non-FRET transfer is likely occurring between the Flo and Rox fluorophores in solution.

Having determined the effect of annealing the two uncleaved top and bottom strands of the reporter DNA, as well as investigating the possible annealing of the "cleaved" single strands, we next turned to the effect of SgrAI binding on the fluorescence signal from the reporter DNA. In **Figure S3A**, the fluorescence emission of 50 nM Flo-26M-2-Rox dsDNA in reaction buffer (without Mg^{2+} and at 25°C) is shown in green (excitation at 495 nm, emission at 520 nm). The addition of 1 μM of SgrAI (gold, **Fig. S3A**) results in ~70% quenching of the Flo emission (after correction for dilution). The binding affinity of SgrAI for this reporter DNA is estimated to be ~2.6 nM[1, 2], and with 1 μM SgrAI, nearly all of the 50 nM reporter DNA should be bound in the SgrAI/DNA complex. Since Mg^{2+} was not present in the reaction solution, the DNA will not be cleaved by SgrAI and the observed fluorescence quenching must therefore be due to SgrAI binding to the reporter DNA. The addition of activator PC DNA at 100 nM and then 1 μM (purple and blue, **Fig. S3A**) has little effect, indicating that filamentation does not result in further quenching of the Flo emission. However, the addition of 10 mM $MgCl_2$ to the reaction (red, **Fig. S3A**) induces a slow unquenching process occurring on a timescale comparable to that seen for the cleavage of this reporter DNA in the DNA cleavage measurements used in this study. Near the end of this process, ~45% of the original Flo emission is recovered (after taking into account the dilution from reagent addition to the cuvette). To understand why 100% of the Flo emission was not recovered, 0.1% SDS was added to the mixture to denature SgrAI and cause release of any bound DNA. Only then was full recovery of the Flo emission was seen (black line, **Fig. S3A**). An independent test verified that

0.1% SDS does not affect the Flo emission itself (**Fig. S3B**). Under the conditions used in **Figure S3A**, all of the reporter DNA is expected to be cleaved at the end of the reaction, hence the residual quenching of the Flo emission before SDS was added indicates that a significant amount of cleaved reporter DNA remains bound to SgrAI at the end of the reactions.

The results of **Figure S3A** suggest that SgrAI binds to the cleaved 26 bp reporter DNA, since full recovery of the fluorescence emission occurred only after addition of SDS (which denatures SgrAI causing release of any bound DNA). To investigate this possibility further, a similar experiment as **Figure S3A** was performed but with the "cleaved" version of the reporter DNA. **Figure S3C s**hows the emission at 520 nm (with excitation at 495 nm) of 50 nM Flo-26M-2-top-cut in reaction buffer without Mg^{2+} and at 25°C (green). The addition of its complement, 26M-2-bot-cut, has little effect (light blue). Rox-26M-2-bot-cut with its complement were not added, since Flo-26M-2-top-cut and its complement, when annealed, can further anneal via their CCGG overhanging base pairs leading to a full "pre-cleaved" primary site which would be capable of binding to SgrAI. The addition of 1 μM SgrAI (gold) to the DNA resulted in slow quenching of about 8% of the initial fluorescence after 2000 sec (after correction for dilution). The addition of PC DNA at 100 nM final concentration (purple) and 1 μM (blue) further quenches the Flo emission, likely by stabilizing the SgrAI/DNA complex and shifting the equilibrium towards binding the cleaved DNA.

We investigated further the idea that the cleaved reporter DNAs may be re-annealing and binding to SgrAI by determining the equilibrium dissociation constants for the cleaved versions of the DNA, and for SgrAI binding to these DNA. The melting temperature of the two half-site duplexes (i.e. Flo-26M-2-top-cut with 26M-2-bot-cut and Rox-26M-2-bot-cut with 26M-2-top-cut) are estimated at 11°C, but binding to SgrAI, and further stabilization of the SgrAI/DNA complex by filament formation may result in pulling the equilibria towards the annealed form of the DNA. In **Figure S4A**, we measured the equilibrium dissociation constant for the annealing of the Flo-labeled half-site (i.e. Flo-26M-2-top-cut and 26M-2-bot-cut) sing fluorescence anisotropy of the fluorescein (excitation at 495 nm and emission at 520 nm). The reactions were performed in 1.5 ml of reaction buffer without $MgCl_2$ and at 25°C. The data were fit to a binding isotherm giving a K_D of 11.3 μM. Such a high

K_D indicates that very little annealing should occur for DNA at 50 nM concentration. Next, the cleaved DNA was held at 50 nM and titrated with SgrAI, in an attempt to directly determine the apparent equilibrium binding constant for SgrAI to the cleaved reporter DNA. Because each cleaved half-site can anneal with any other cleaved half-site via its CCGG overhang, we used only the left half-site consisting of Flo-26M-2-top-cut and its complement. Concentrations of SgrAI above 2.2 μM were not possible, and the plots of **Fig. S4B-C** indicate that saturation of the cleaved DNA by SgrAI was not obtained even at the highest concentration of SgrAI tested. Nevertheless, fitting of the data to an equilibrium binding equation gave a K_D of 4.2 μM, although we consider this to be a lower estimate of the true value. This K_D gives a forward binding rate constant for binding of SgrAI to the cleaved DNA of $\leq 1 \times 10^{-6}$ $nM^{-1}s^{-1}$ (if the dissociation rate constant is 0.4 s^{-1}, a value used in the data fitting, **Table 1**). Of course, the addition of PC DNA and induction of filamentation would pull the linked equilibria towards greater binding of the cleaved DNA by SgrAI, as was seen in the test of **Fig. S3A**.

To see the effect of including a nonzero forward binding rate constant for SgrAI and cleaved DNA (which was assumed to be zero in previous modeling with shorter reporter DNAs[3]), we conducted a series of 1D and 2D Fitspace calculations in Kintek GKE. In these calculations, χ^2 is used to determine the quality of the fit of the predictions by the model to the experimental data[4]. Fitspace systematically tests values for a single rate constant (1D), or two rate constants (2D), allowing global fitting of all experimental data by adjusting relevant scale factors and any rate constants not held constant, and then determines the χ^2 of the best fit with each tested rate constant. In **Figure S5A**, the forward rate constant of binding the cleaved DNA to SgrAI (i.e. an apparent forward rate constant encompassing all equilibria including strand dissociation and annealing of the reporter DNA) was initially set to a value of 1×10^{-4} nM^{-1} s^{-1} but allowed to vary between 1×10^{-6} and 1.0 nM^{-1} s^{-1} as the rate constant for conversion of the H state of the SgrAI/DNA complex to the L state was varied between 1 and 11 s^{-1} (note that the reverse rate constant for this interconversion was held constant at 1 s^{-1}, hence the parameter varied is the ratio of the two rate constants, the L/H ratio). Best fits give the lowest χ^2, hence the peak in the plot of χ^2_{min}/χ^2 in **Figure S5A** shows that the best value for the L/H ratio is 5.9. Values giving χ^2_{min}/χ^2 within 90% of the best χ^2_{min}/χ^2 are 5.4 to 6.6 for the L/H ratio. **Figure S5B** shows the result on χ^2_{min}/χ^2 when

varying the apparent rate constant for binding of SgrAI to the cleaved DNA between $1x10^{-6}$ and 1.0 nM^{-1} s^{-1}, and the L/H ratio allowed to be adjusted to any value between 1 and 11 to give the lowest χ^2. No peak in χ^2_{min}/χ^2 is observed, indicating that the quality of the fit is insensitive to this parameter within this range of values and that the experimental data must not therefore contain information on this equilibrium. **Figure S5C** show the results of a 2D Fitspace calculation using these same two parameters as variables in global fitting. The 2D grid used is 10x10 and the plot contains the results of 100 fitting calculations, one for each value of the two varied rate constants. The plot shows that the fits are highly sensitive to the value of the L/H ratio (varied on the x-axes, with an optimum ~5.4-6.6), but completely insensitive to the rate constant of binding to the cleaved DNA (varied on the y axes).

In summary, these data show that little FRET occurs within the reporter DNA and that most of the quenching and unquenching of the Flo emission seen in DNA cleavage reactions is a result of SgrAI binding and unbinding to the reporter DNA. The data also show that although re-annealing of the cleaved DNA is highly unfavorable at the concentrations used in the reactions, the binding to SgrAI shifts the equilibrium towards annealing to a small degree. The apparent dissociation constant of SgrAI and the cleaved reporter DNA is in the micromolar range, but because SgrAI is at 1 μM in the reactions, some binding to the cleaved DNA is observed. We reason that because the steps involved (dissociation and rebinding of SgrAI with the annealed cleaved DNA and dissociation/annealing of cleaved DNA) are fast (i.e. the forward and reverse rate constants are large), and the concentration of the cleaved DNA is low (50 nM), the *percentage* of cleaved DNA bound to SgrAI does not change as the DNA is cleaved and released during the course of the reaction. As a result, this factor does not influence the observed *rate* of release of the cleaved DNA by SgrAI (only the percentage of fluorescence recovered) and has no effect on global model fitting as shown by the Fitspace tests in **Figure S5**.

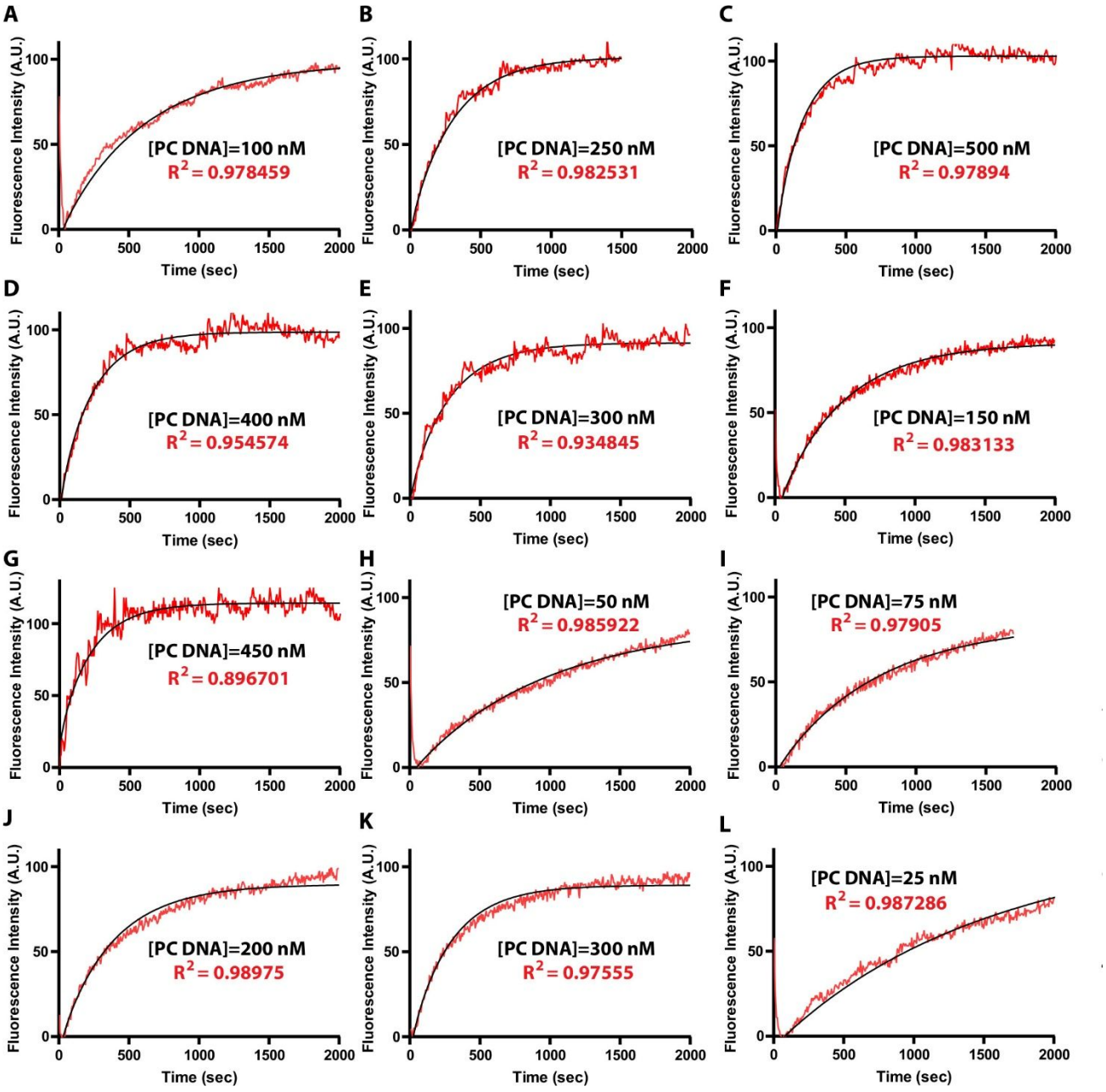


Figure S6. Fluorescence unquenching and simulated curves from global data fitting to the Kintek model. **A-L.** Normalized fluorescence measured for experiments 3-9, 13-17 (red) and corresponding simulated data from Kintek modeling (black). The concentrator of activator PC DNA, and the R^2 of the fit of the simulated curve to the experimental data is given for each plot.

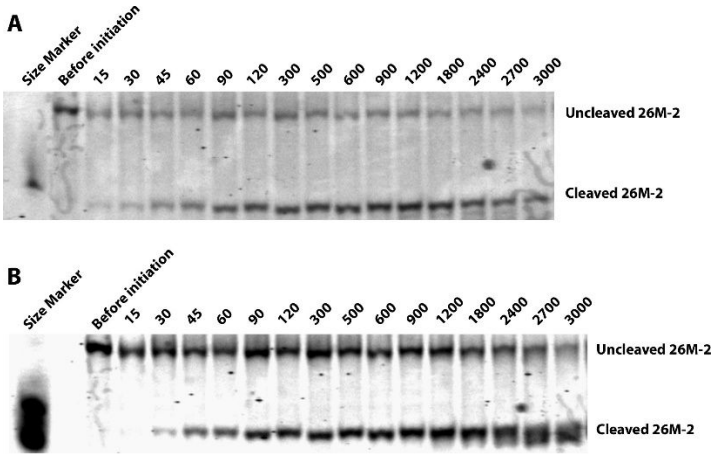


Figure S7. Representative scans of denaturing PAGE gels used to quantitate DNA cleavage. A. Gel scan for the Flo labeled strand in Experiment 10 which contained 1 μM SgrAI, 50 nM Flo-26M-2-Rox, and 250 nM activator PC DNA. **B.** As in A, but gel scan for the Rox labeled strand.

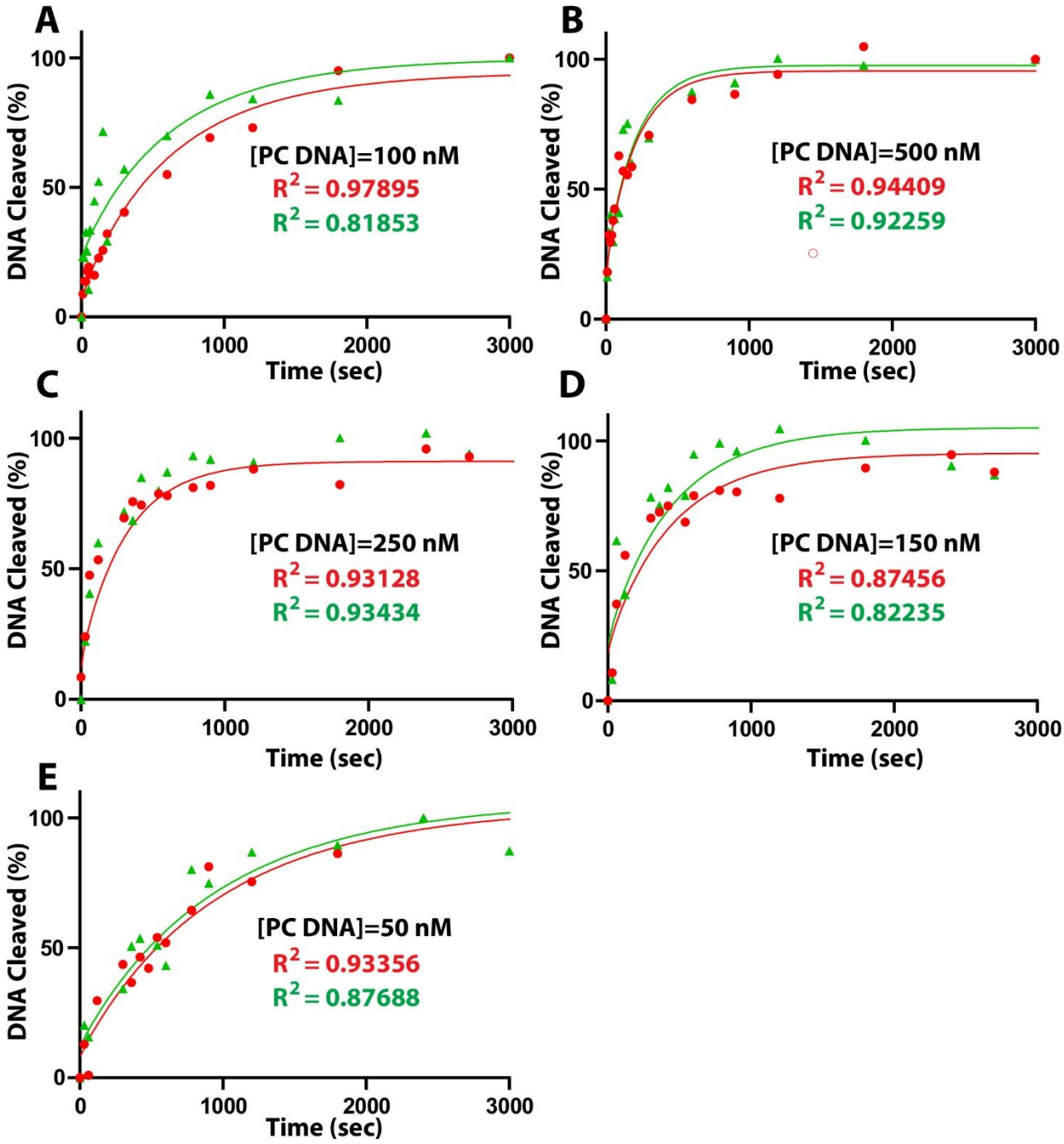


Figure S8. Experimental DNA cleavage data measured using denaturing gel electrophoresis and simulated curves from global data fitting. A-E. Measured percentage of DNA cleaved at indicated times after initiation for experiments 1, 2, 10-12 and corresponding simulated data from Kintek modeling. The x-axes represent the time from reaction initiation in seconds. Red filled circles and green squares are measured data at indicated times for the Rox and Flo labeled strands, respectively. Lines represent simulated data using Kintek

Global Kinetic Explorer. The concentration of activator PC DNA and R^2 measuring the agreement between the simulated and experimental data is also given for each plot.

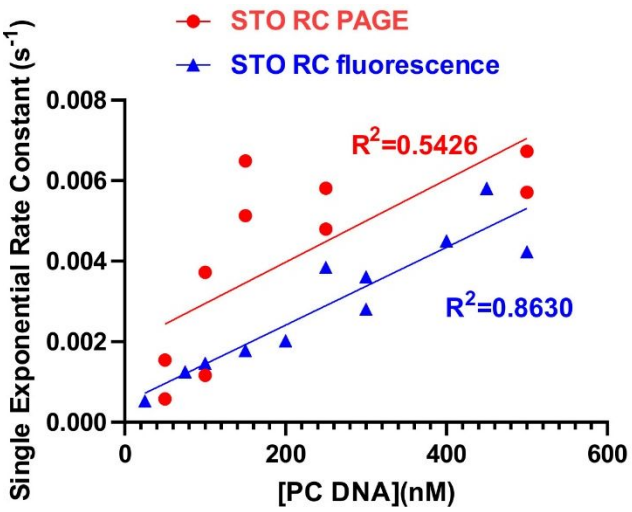


Figure S9. Rate constants for each data set determined using a single exponential equation. Rate constants determined by PAGE are shown as red circles, and those determined by fluorescence are shown as blue triangles. Lines were fit to data to determine an R^2 for the linear dependence of the rate constant on the concentration of PC DNA for data measured by each method. The rate constants measured with PAGE are faster in general than those measured using fluorescence. The data from PAGE is noisier than that from fluorescence as well.

Comparison of analytical fits of PAGE and fluorescence data

A single exponential equation:

$$Percentage\ of\ cleave\ reporter\ DNA = baseline + scaling\ factor\ \times\ e^{-kt}$$

or

$$Fluorescence\ signal = baseline + scaling\ factor\ \times\ e^{-kt}$$

were fit to each experimental data set to determine a rate constant k (given in s^{-1}). These data are also shown in **Table S1**, along with the goodness of fit (R^2) measuring the agreement between the fit and experimental data. Some data sets were fit better by this function than the Kintek simulated curves (Experiments 1, 7, 11, 12, 13, 14, and 16), while others were fit better by Kintek (Experiments 2, 3, 5, 8, 13, 15, and 17), and some fit equally

well by both (Experiments 4, 6, 9, 11, and 16). We found no systematic trend in activator DNA concentration (in terms of PC DNA concentration) for which analyses fit better, with one exception. Experiments with lower concentrations of activator, and measured via fluorescence quenching,were more likely to be fit better with Kintek simulations than with a single exponential function. This is likely due to the fact that with lower PC DNA concentrations, the reaction progress is multiphasic, beginning with a slow phase, transitioning to a faster phase, and ending with a slow phase. The simulations created by the Kintek model include the influence of multiple phases, but a single exponential function includes only one. Both approaches fit the experimental data equally well when the activator PC DNA concentration are high, because only a single fast phase dominates the reaction. It should be noted that the Kintek modeling uses multiple steps each with forward and reverse rate constants to fit the experimental data, however the Kintek modeling also attempts to find the best fit of these rate constants for all 17 data sets simultaneously, whereas the fits with single exponential functions were fit to each data set independently.

Figure S9 plots the single exponential rate constants from fits of the experimental data used herein. Data measured using PAGE (red circles, **Fig. S9**), and those measured from fluorescence unquenching (blue triangles, **Fig. S9**) were plotted vs. the concentration of activator PC DNA used in each reaction. A systematic trend is observed where these rate constants increase as the PC DNA (i.e. the activator DNA) concentration is increased. The data measured using fluorescence (blue triangles, **Fig. S9**) shows a better linear relationship as measured by the R^2 of a best fit line (0.8630) compared to that for the PAGE data (red circles, R^2=0.5426). The lower R^2 of the PAGE data by this analysis is likely due to the noisier data in general (**Fig. S8)**, which resulted from high background in scans of the fluorescence in the gels.

An interesting observation shown in **Figure S9** is that the single exponential rate constants are higher for PAGE data than for the fluorescence data, with the exception of those measured at low PC DNA concentration. In reactions with lower PC DNA concentrations, the association of SgrAI DBD into filaments is rate limiting, which impacts both types of measurements equally (i.e. the measurement of DNA cleavage by PAGE and the measurement of cleaved DNA release using fluorescence unquenching). In reactions with higher concentrations of PC DNA, later steps in the reaction pathway influence the observed rates to a much greater degree, such as

the filament dissociation step following DNA cleavage. Data measured with PAGE does not include this step, hence data measured by PAGE exhibit rate constants higher than those measured by fluorescence (compare red circles to blue triangles, **Fig. S9**).

Table S2. Kintek Global Kinetic Explorer Equations

The model named 4EO from Park, et al., 2018[3] was used in global model fitting. Note that this model includes the following characteristics:

- All reactions are reversible, with one exception (DNA cleavage)
- PC is the pre-cleaved primary site analogous to one-half of a full activator DNA. Two copies assemble into PCDS, which is the activator DNA.
- PCDS binds to SgrAI to form "F"
- M26_2 is the reporter DNA Flo-26M-2-Rox, which binds to SgrAI to form "RT"
- "RT" is also the L state conformation of SgrAI bound to the reporter DNA
- RT is in equilibrium with "R", the H state conformation of SgrAI bound to the reporter DNA, this is the new step added in the current work.
- Only R can bind to SgrAI bound to activator DNA (i.e. "F") to form or join filaments
- R cannot bind another R, only an F
- Filaments form by the addition of R or F at either end, with the rules that F can bind F, and R can bind F, but R cannot bind R
- Within the filament, R cleaves DNA to form "RX"
- Filaments with cleaved DNA (i.e. with "RX") dissociate to release the RX. This reaction is reversible.
- Dissociation of F or RX occurs only from a filament end
- Filaments up to 4 copies of R or F can form.
- RX can dissociate into SgrAI and cleaved DNA ("X"). This reaction is considered reversible, though the reverse (i.e. binding) rate constant is not determined by fitting to the data.
- SgrAI bound to the reporter DNA, but not in a filament, can cleave DNA when in the R state

PC + PC = PCDS

SgrAI + PCDS = F

SgrAI + M26_2 = RT

RT = R

F + F = FF

F + R = FR

R + F = RF

F + FR = FFR

FF + R = FFR

F + RF = FRF

FR + F = FRF

RF + F = RFF

R + FF = RFF

R + FR = RFR

RF + R = RFR

FF + F = FFF

F + FF = FFF

FFF + F = FFFF

F + FFF = FFFF

FFF + R = FFFR

F + FFR = FFFR

FFR + F = FFRF

F + FRF = FFRF

FRF + F = FRFF

F + RFF = FRFF

RFF + F = RFFF

R + FFF = RFFF

R + FFR = RFFR

RFF + R = RFFR

R + FRF = RFRF

RFR + F = RFRF

F + RFR = FRFR

FRF + R = FRFR

FR = FRX

RF = RXF

FFR = FFRX

FRF = FRXF

RFF = RXFF

RFR = RXFRX

FFFR = FFFRX

FFRF = FFRXF

FRFF = FRXFF

RFFF = RXFFF

RFRF = RXFRXF

FRFR = FRXFRX

RFFR = RXFFRX

F + RX = FRX

RX + F = RXF

F + FRX = FFRX

FF + RX = FFRX

F + RXF = FRXF

FRX + F = FRXF

RX + FF = RXFF

RXF + F = RXFF

RX + FRX = RXFRX

RXF + RX = RXFRX

F + FFRX = FFFRX

FFF + RX = FFFRX

F + RXFRX = FRXFRX

FRXF + RX = FRXFRX

RX + FFRX = RXFFRX

RXFF + RX = RXFFRX

RX + FRXF = RXFRXF

RXFRX + F = RXFRXF

F + FRXF = FFRXF

FFRX + F = FFRXF

F + RXFF = FRXFF

FRXF + F = FRXFF

RX + FFF = RXFFF

RXFF + F = RXFFF

RX = SgrAI + X

R = RX

Kintek equations for simulating experimental data:

PAGE data: BL+SF*X

Fluorescence dequenching data:

BL+(SF)*(X+RX+FRX+RXF+FFRX+FRXF+RXFF+FFFRX+FFRXF+FRXFF+RXFFF+2*(RXFRX+FRXF

RX+RXFRXF+RXFFRX))

REFERENCES

[1] Park, C. K., Stiteler, A. P., Shah, S., Ghare, M. I., Bitinaite, J., and Horton, N. C. (2010) Activation of DNA cleavage by oligomerization of DNA-bound SgrAI, *Biochemistry 49*, 8818-8830.

[2] Shah, S., Sanchez, J., Stewart, A., Piperakis, M. M., Cosstick, R., Nichols, C., Park, C. K., Ma, X., Wysocki, V., Bitinaite, J., and Horton, N. C. (2015) Probing the run-on oligomer of activated SgrAI bound to DNA, *PLoS One 10*, e0124783.

[3] Park, C. K., Sanchez, J. L., Barahona, C., Basantes, L. E., Sanchez, J., Hernandez, C., and Horton, N. C. (2018) The run-on oligomer filament enzyme mechanism of SgrAI: Part 2. Kinetic modeling of the full DNA cleavage pathway, *J Biol Chem 293*, 14599-14615.

[4] Johnson, K. A., Simpson, Z. B., and Blom, T. (2009) FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data, *Anal Biochem 387*, 30-41.